PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 31 August 15, 1945 Number 8
*Copyright 1945 by the National Academy of Sciences*

# *X-RAY INDUCED MUTANT STRAINS OF ESCHERICHIA COLI**

BY E. L. TATUM

SCHOOL OF BIOLOGICAL SCIENCES, STANFORD UNIVERSITY

Communicated June 20, 1945

In a previous paper the production by means of x-ray treatment of a number of growth-factors requiring mutant strains of *Escherichia coli* and of *Acetobacter melanogenum* was described.[1] These results were interpreted as indicating an analogy with the induction by irradiation of true gene mutations in *Neurospora* with similar biochemical changes.[2, 3, 4]. Comparable changes have been induced in other strains of *E. coli*.[5, 6] The present work was undertaken with the aims of producing more "biochemical" mutant strains of *E. coli* for biochemical study, of providing convincing evidence of derivation of these strains and of comparing the mutation rates of irradiated and control cultures. A considerable number of new mutant strains have been obtained after further irradiation of two biochemical mutant strains previously described.[1] These new strains require growth-factors in addition to those required by the parent strains. The double requirements of the new strains thus provide reasonable certainty of their derivation from their parent strains. The irradiation significantly raised the mutation rate from one in 2000 isolations to 16 in the same number of isolations. Furthermore, the biochemical specificities of those mutant strains which have been analyzed are consistent with the view that each differs from the original stock in only one biochemical reaction. The evidence therefore supports the view that treatment with x-rays produces heritable defects in synthetic reactions in bacteria, biologically and biochemically analogous to those resulting from single gene mutations in *Neurospora*. This is presumptive evidence for the existence of genes in bacteria, perhaps contained in the nuclear structures which have been observed in a number of bacteria.[7]

*Experimental.*—Two previously obtained mutant strains of *E. coli* have been used in these investigations. Strain 58 is characterized by requiring biotin, and strain 679, threonine.[1] These mutant strains were used, rather than the wild type strain from which both were derived, so that their char-

acteristic growth-factor requirements would serve as biochemical "markers." Both of these strains had been maintained for over a year on complete agar medium (containing yeast extract, peptone and glucose) with no changes in their minimum requirements. Twenty-four-hour-old cultures grown in complete liquid medium with shaking were irradiated with approximately 180,000 r units during a period of 45 minutes, as previously described.[1] After a further incubation period of four hours, samples of the irradiated suspensions were plated out in complete medium, incubated for several days and then isolated colonies transferred to complete agar medium. These isolated strains were then tested for growth in the liquid minimal medium[1] which for testing new strains derived from strain 58 was supplemented with 0.1 $\mu$g. biotin per 100 ml., and for testing those derived from strain 679 was supplemented with 10 mg. *dl*-threonine per 100 ml. Cultures which consistently failed to grow in the minimal media were then tested for their specific additional growth requirements. The data in table 1 summarize the results obtained. A mutation rate of around 1 per cent was obtained with both irradiated strains, as contrasted with a rate of 0.05 with the unirradiated culture, apparently a significant increase on irradiation to approximately the same value previously obtained with *Acetobacter melanogenum*, but somewhat higher than that obtained with the original strain of *E. coli* (K-12).[1]

TABLE 1

EFFECT OF X-RAY TREATMENT ON PRODUCTION OF MUTANT STRAINS OF *E. coli*

| STRAIN OF ORIGIN[1] | X-RAY TREATMENT | VIABLE CELLS, %* | NUMBER OF COLONIES ISOLATED | DESIGNATION OF STRAINS | NUMBER OF MUTANT STRAINS |
|---|---|---|---|---|---|
| 58 | 180,000 r | 0.001 | 941 | 58-1 to 58-941 | 9 |
| 679 | 180,000 r | 0.01 | 800 | 679-1 to 679-800 | 7 |
| 58 | None | 100.0† | 1902 | 58-1001 to 58-2902 | 1 |

* Calculated from direct microscopic and plate counts.

† This culture was 72 hours old at the time of plating.

Of the 17 mutants obtained all but three have been found to require the addition of only one known growth factor for normal growth. Table 2 gives the substance required by each of these strains, together with the concentration which is required for $^1/_2$ maximal growth from turbidity measurements. The growth response of each of these strains is a function of the concentration of the growth factor supplied. Most of the mutant strains showed amino acid deficiencies, but three strains required thiamin, and three failed to grow on hydrolyzed casein or on a mixture of vitamins, but grew well on yeast extract. The only apparent duplications of requirements were those of the proline requiring strains 679-183 and 679-440, and of the thiamin requiring strains 58-593 and 58-610. In all other cases, either different substances were required or the strain of origin was different,

so that of the 17 mutant strains isolated, 14 were definitely of independent origin.

After the required substance was determined, each strain derived from strain 58 was re-tested for its requirement for biotin, and each one derived from strain 679 re-tested for its need for threonine. In every case the original requirement persisted in addition to the second induced requirement. The possession of these double requirements is definite evidence of the origin of the new strains. The biochemical specificities of some of these mutant strains have also been tested. Table 2 also lists the biochemically related substances tested for activity on each mutant strain.

TABLE 2

DESCRIPTION OF MUTANT STRAINS OF *E. coli*

| STRAIN NUMBER | SUBSTANCE REQUIRED | CONCENTRATION GIVING 1/2 MAXIMAL GROWTH IN 72 HOURS* PER 10 ML. | RELATED SUBSTANCES TESTED ACTIVE | INACTIVE |
|---|---|---|---|---|
| 58-161 | Methionine | 0.01 mg. | Homocystine | $Na_2S$, cystine |
| 58-278 | Phenylalanine | 0.016 mg. | ..... | Tyrosine |
| 58-309 | Cystine | 0.16 mg. | ..... | $Na_2S$, methionine ± serine, homocystine ± serine |
| 58-336 | Isoleucine | 0.04 mg. | Hydroxy- and keto-acid analogues | ..... |
| 58-580 | Thiamin | 0.004 μg. | ..... | Thiazole + pyrimidine |
| 58-593 | Thiamin | 0.003 μg. | Thiazole | Pyrimidine |
| 58-610 | Thiamin | 0.003 μg. | Thiazole | Pyrimidine |
| 58-741 | Histidine | 0.003 mg. | ..... | ..... |
| 679-183 | Proline | 0.02 mg. | ..... | Ornithine, arginine, glutamic acid, hydroxyproline |
| 679-440 | Proline | 0.03 mg. | ..... | Ornithine, arginine, glutamic acid, hydroxyproline |
| 679-662 | Glutamic acid or proline† | 0.05 mg.<br>0.05 mg. | ..... | Ornithine, arginine, hydroxyproline |
| 679-680 | Leucine | 0.03 mg. | Hydroxy-acid analogue | ..... |
| 58-2651‡ | Proline | 0.03 mg. | ..... | Ornithine, arginine, glutamic acid, hydroxyproline |
| 58-178<br>679-447<br>679-455 | Yeast extract | ..... | ..... | Vitamins, hydrolyzed casein |

* From growth curves obtained with an Evelyn photoelectric colorimeter.

† The growth of this strain on proline is slower than on glutamic acid, although quantitatively similar responses are obtained in 72 hours.

‡ Obtained from unirradiated material.

Strains 58-336 and 679-680, which require isoleucine and leucine respectively, can grow with the corresponding hydroxy-acid analogues, and in the one case tested (strain 58-336) also on the keto-acid analogue. Presumably the syntheses of these amino acids are blocked at stages in the formation of the carbon skeletons prior to the formation of the keto acids, which can be aminated to form the amino acids. The methionine requiring strain 58-161 presumably cannot form methionine from reduced sulfur ($Na_2S$) or from cystine, but can do so to a limited extent from homocystine, a process which would require the methylation of homocysteine. The cystine requirement of strain 58-309 is extremely specific. Apparently cystine cannot be formed from $Na_2S$, nor from methionine or homocystine even in the presence of added serine. If the normal synthesis of cysteine in *E. coli* involves the transfer of the —SH group from homocysteine to serine by the reaction studied by Binkley and du Vigneaud,[8] this strain must be unable to transfer the sulfhydryl group to serine.

None of the three strains which respond to proline use ornithine, arginine or hydroxy-proline. However, one can grow on either proline or on glutamic acid. Proline and glutamic acid are apparently interconvertible in strain 679-662, but not in the other two strains. If the failure of synthesis of proline in each of these strains results from the failure of a single biochemical reaction, these facts suggest that ornithine is not involved in the synthesis of proline by *E. coli*, in which glutamic acid may be the immediate precursor of proline.[9]

There are two types of thiamin requiring strains, both analogous to known mutant strains of *N. crassa*.[4, 10] The two representatives of one type, strains 58-593 and 58-610, apparently are unable to synthesize the vitamin thiazole, but can form the pyrimidine component and couple the two components. Strain 58-580, however, cannot carry out this coupling, since it requires intact thiamin.

*Summary.*—Following a second x-ray treatment of two mutant strains of *Escherichia coli* 16 new mutant strains have been obtained from 1741 single colony isolations. One mutant strain was obtained from 1902 isolations from unirradiated material.

Fourteen of these strains have been found to require single known growth factors in addition to the requirement of the strain from which they were derived for biotin or for threonine. The growth factors required by these strains include proline, glutamic acid or proline, leucine, isoleucine, phenylalanine, methionine, cystine, histidine, thiamin and vitamin thiazole. Three strains require unidentified substances present in yeast extract.

These results are consistent with the conclusion that growth-factor requirements in bacteria result from heritable changes analogous to true gene mutations.

* Work supported in part by a grant from the Rockefeller Foundation. The author wishes to thank Dr. David Bonner for supplying the hydroxy- and keto-acid analogues of leucine and isoleucine, and Miss Esther M. Zimmer for technical assistance in carrying out this investigation.

[1] Gray, C. H., and Tatum, E. L., these PROCEEDINGS, **30,** 404 (1944).

[2] Beadle, G. W., and Tatum, E. L., *Ibid.*, **27,** 499 (1941).

[3] Horowitz, N. H., Bonner, D., Mitchell, H. K., Tatum, E. L., and Beadle, G. W., *Amer. Nat.* (in press).

[4] Tatum, E. L., and Beadle, G. W., *Annals Mo. Botanical Garden,* **32,** 125 (1945).

[5] Roepke, R. R., Libby, R. L., and Small, M. H., *Jour. Bact.,* **48,** 401 (1944).

[6] Anderson, E. H., these PROCEEDINGS, **30,** 397 (1944).

[7] For descriptions and for further references see Robinow, C. F., *Jour. Hygiene,* **43,** 413 (1944).

[8] Binkley, F., and du Vigneaud, V., *Jour. Biol. Chem.,* **144,** 507 (1942).

[9] For a discussion of these relationships and for further references, see Shemin, D., and Rittenberg, D., *Jour. Biol. Chem.,* **158,** 71 (1945).

[10] Tatum, E. L., and Bell, T. T. (in preparation).

---

## *CARBON DIOXIDE UTILIZATION IN THE SYNTHESIS OF ACETIC ACID BY CLOSTRIDIUM THERMOACETICUM*

BY H. A. BARKER AND M. D. KAMEN*

DIVISION OF PLANT NUTRITION, UNIVERSITY OF CALIFORNIA

Communicated June 20, 1945

In most carbohydrate fermentations the appearance of acetic acid or other $C_2$ products is accompanied by at least an equimolar quantity of carbon dioxide or other $C_1$ product. The carbohydrate fermentations caused by *Clostridium thermoaceticum* are a noteworthy exception to this generalization since the only product other than cell material is acetic acid.[1, 2] Specifically, glucose and xylose are acted upon by this organism to yield, respectively, about 2.65 and 2.25 moles of acetic acid per mole of carbohydrate decomposed. When pyruvate is the substrate, both acetic acid and carbon dioxide are formed but the yield of the latter is low, corresponding to about 0.5 mole per mole of pyruvate.

To account for the high yield of acetic acid and the low yield of carbon dioxide from all three substrates, it has been postulated that *Clostridium thermoaceticum*, like *Clostridium aceticum*[3, 4] and *Clostridium* acidi-urici,[5] uses carbon dioxide as an oxidizing agent in such a way that it is condensed with a second molecule and reduced to acetic acid. If this occurs in the decomposition of glucose, for example, at least two alternative mechanisms, illustrated by the following sets of equations, are possible.

I. $C_6H_{12}O_6 + 2H_2O = 2CH_3COOH + 2CO_2 + 8H$ (1)

$8H + 2CO_2 = CH_3COOH + 2H_2O$ (2)

II. $C_6H_{12}O_6 + 6H_2O = 6CO_2 + 24H$ (3)

$6CO_2 + 24H = 3CH_3COOH$ (4)

For either scheme the over-all reaction is

$$C_6H_{12}O_6 = 3CH_3COOH. \quad (5)$$

Scheme I represents the fermentation as a partial oxidation of glucose to acetic acid and carbon dioxide (equation 1) accompanied by the reduction of 2 moles of carbon dioxide to acetic acid (equation 2). The oxidation of glucose presumably follows the usual glycolytic mechanism involving pyruvic acid. Scheme II represents the fermentation as a complete oxidation of glucose (equation 3) accompanied by the reduction of 6 moles of carbon dioxide (equation 4) to acetic acid. Although Scheme II appears less likely to represent the true course of the fermentation than does Scheme I, it must nevertheless be considered as a possibility.

In this paper experiments will be described which provide direct evidence for the conversion of carbon dioxide to acetic acid and for the decomposition of glucose in accordance with Scheme I. This evidence has been obtained[6] by the use of carbon dioxide labeled with the long-lived carbon isotope, $C^{14}$.

*Experimental Methods.*—For the detection of radiation emitted by $C^{14}$ a Geiger counter of "bell-jar" type with a thin mica window was used. The mica window had a thickness corresponding to about 3.5 mg. per cm.$^2$ and an area of approximately 25 cm.$^2$ It was supported by a brass grid cut to allow maximal passage of the radiation. Samples consisting of barium carbonate or barium acetate were spread thinly and evenly on duraluminum discs and dried at 100°C. These were placed about 1 mm. below the window in a standard fixed position. Variations in sensitivity of the counter were corrected for by reference to a standard sample of barium carbonate prepared in the same way. Corrections for self-absorption, necessitated by the softness of the beta rays (0.15 M. E. V.) emitted, were determined from a curve constructed from counts on standard samples of known activity and varying thickness.† The long half-life of $C^{14}$ ($2.5 \pm 1 \times 10^4$ yrs.)[7] obviated corrections for decay. The counting circuit was of a conventional type and requires no description.

The bacteria were grown in a medium of the following composition in g. per 100 ml.: glucose 0.15–0.7, tryptone 0.5, yeast extract 0.5, pH 6.6, phosphate buffer 0.9, $(NH_4)_2SO_4$ 0.05, $MgSO_4 \cdot 7H_2O$ 0.01, sodium thioglycollate 0.05, $Na_2C^*O_3$ about 0.2 (* indicates carbon labeled with $C^{14}$). The phosphate buffer and labeled sodium carbonate were added after autoclaving. In a typical experiment the total volume of medium was about 12 ml. Oxygen was removed by means of an Oxsorbent seal and

the culture tube was closed with a ground-glass stopper to prevent loss of carbon dioxide. After incubating for 3 to 6 days at 55°C. cultures were analyzed and the $C^{14}$ content of the carbon dioxide and the fermentation products was determined.

*Results.*—In table 1 are presented data from an experiment such as that described above. The glucose concentration was 0.7 g. per 100 ml.

TABLE 1

THE FERMENTATION OF GLUCOSE IN THE PRESENCE OF $C^*O_2$
Experiment 1

| SUBSTANCE | MG./10 ML. | CTS./MIN./MG. | TOTAL CTS./MIN. |
|---|---|---|---|
| Glucose fermented | 54.0 | ... | ..... |
| Initial carbon dioxide as $BaCO_3$ | (22.3) | 117 | 2610 ± 50 |
| Final carbon dioxide as $BaCO_3$ | 22.3 | 5.7 | 128 ± 20 |
| Acetic acid formed as $BaAc_2$ | 101.6 | 19.9 | 2020 ± 40 |
| Cell material (residue from T. C. A. extraction) | 2.5 | 12.8 | 32 ± 4 |
| Trichloracetic acid extract of cells | 4.5 | 1.5 | 7 ± 4 |
| Non-volatile cell-free material | ... | ... | 96 ± 30 |
| Total $C^{14}$ activity in products | | | 2155 |
| $C^{14}$ recovered in per cent | | | 2155/2482 × 100 = 87 |

It can be seen that most (94%) of the $C^{14}$ disappeared from the added carbon dioxide during the fermentation and a large part of it (81%) was recovered in acetic acid. Adequate evidence that the $C^{14}$ was actually present in acetic acid rather than some associated compound was provided by distilling the volatile acid, using a modified Duclaux method, and determining the specific activity of the barium acetate derived from successive distillation fractions. The specific activity was found to be the same in all fractions within the limits of experimental accuracy (table 2).

TABLE 2

SPECIFIC ACTIVITIES OF DUCLAUX FRACTIONS OF ACETIC ACID
Experiment 1 (total volume 110 ml.)

| DUCLAUX FRACTION | CTS./MIN./MG. $BaAc_2$ |
|---|---|
| 0– 40 ml. | 17.9 |
| 40– 80 ml. | 18.9 |
| 80–110 ml. | 18.7 |

A small part of the $C^{14}$ was also present in the cells after removal of acetic acid and carbon dioxide, and in the non-volatile fraction of the medium after removal of the cells. It is noteworthy that the quantity of $C^{14}$ per mg. of dry cells was of the same order of magnitude as in the barium acetate. It seems reasonable to conclude that a considerable part of the cell material was synthesized from carbon dioxide.

Only about one-fifth of the $C^{14}$ in the cells could be extracted with 4% trichloracetic acid; the remainder must have been present in proteins,

lipoids and other acid-insoluble forms. The small yield of acid-extractable $C^{14}$ was due possibly to the relatively long incubation which was continued for several days after growth ceased. The bacterial cells may have autolyzed, releasing soluble constituents. This could account for the considerable amount of non-volatile $C^{14}$ present in the medium after removal of the bacteria. The actual weight of this fraction could not be estimated because of the large amount of inorganic salt present; for the same reason the $C^{14}$ content of this fraction could be only roughly determined.

To determine the distribution of $C^{14}$ in the acetic acid, its barium salt was first decarboxylated according to the equation

$$(CH_3COO)_2Ba = CH_3COCH_3 + BaCO_3.$$

In control experiments with synthetic $CH_3C^*OOH$, the barium carbonate was shown to originate from the carboxyl carbon of acetic acid (table 3, column 2). Hence, this method was used to estimate the $C^{14}$ content of the carboxyl carbon. To determine the $C^{14}$ content of the methyl group, the acetone derived from the decarboxylation was oxidized with alkaline iodine to iodoform and acetic acid. The iodoform, which was shown to originate from the methyl group of acetic acid (table 3, column 2), was oxidized to carbon dioxide and converted to barium carbonate for counting $C^{14}$.

TABLE 3

DISTRIBUTION OF $C^{14}$ IN SYNTHETIC AND FERMENTATION ACETIC ACID

(The figures give the percentage of the total $C^{14}$ in each atom)

| CARBON ATOM | SYNTHETIC $CH_3C^*OOH$ | FERMENTATION ACETIC ACID EXPT. 1 | EXPT. 3 |
|---|---|---|---|
| Methyl | 0 ± 3 | 49 ± 2 | 40 ± 2 |
| Carboxyl | 100 ± 2 | 51 ± 2 | 60 ± 2 |

Data from two experiments are given in table 3, columns 3 and 4. It can be seen that in experiment 1 the $C^{14}$ in the acetic acid produced by fermentation was almost equally distributed between the methyl and carboxyl groups as would be expected if a total synthesis of acetic acid from carbon dioxide had occurred. In experiment 3, a different result was obtained; about 50% more $C^{14}$ was present in the carboxyl than in the methyl group. The difference in $C^{14}$ distribution between the two experiments appears to be significant but we have no explanation for it. Further work will be necessary to elucidate the relation between the experimental conditions and the isotope distribution in acetic acid. There is, however, a possible explanation for the fact that in experiment 3 the carboxyl group had a higher $C^{14}$ content than the methyl group. The synthesis of acetic acid from carbon dioxide by *Cl. thermoaceticum* may occur in two ways, one involving the fixation of carbon dioxide in both the methyl and carboxyl groups, the other involving only fixation in the carboxyl group.

The relative rates of the two processes would then determine the distribution of the isotope. A fixation of carbon dioxide exclusively in the carboxyl group of acetic acid has been observed by Slade, *et al.*,[8] using *Aerobacter indologenes* and *Clostridium welchii*.

Having established the conversion of carbon dioxide to acetic acid, it seemed desirable to find out how much carbon dioxide was produced as an intermediate in the glucose fermentation. This can be calculated from a knowledge of the $C^{14}$ contents of the initial and final carbon dioxide if it is assumed that the formation of carbon dioxide from sugar is the only process which causes dilution of the labeled carbon dioxide. One specific reaction which must not occur is an exchange of $C^{14}$ between carbon dioxide and preformed acetic acid.

The evidence against a reversible exchange of $C^{14}$ between carbon dioxide and acetic acid is of two types. First, a wide variation in the molar ratio of the $C^{14}$ contents of acetic acid and carbon dioxide was observed in different experiments. The range was from 0.95 (experiment 2, table 4) to 2.26 (experiment 1, table 1). A rapid exchange of $C^{14}$ between the two compounds should result in a constant ratio. Secondly, an experiment was performed in which glucose was fermented in the presence of acetate, labeled in both the methyl and carboxyl positions, and ordinary carbon dioxide. At the beginning of the experiment 0.215 mM of acetic acid giving 726 ± 15 cts./min. was present per 10 ml. of medium. After the fermentation 0.385 mM of acetic acid giving 706 ± 15 cts./min. was recovered. The final carbon dioxide (0.049 mM) gave only 9.8 ± 1 cts./min., or about 1% of the $C^{14}$ initially added as acetate. From this result it must be concluded that whereas a conversion of acetic acid carbon to carbon dioxide does occur the rate of the reaction is so slow as to be negligible under the conditions of these experiments.

To calculate the intermediate carbon dioxide production we shall assume then that the only pertinent reactions are the following:

$$(a) \quad \text{glucose } (C^{12}) \rightarrow CO_2\ (C^{12})$$

$$(b) \quad CO_2\ (C^{12} + C^{14}) \rightarrow \text{acetic acid } (C^{12} + C^{14}).$$

The carbon dioxide will always consist of a mixture of $C^{14}O_2$ and $C^{12}O_2$. Let $x$ be the quantity of $C^{14}O_2$ per unit volume at any time during the fermentation and let $x_0$ be the initial, and $x_f$ the final $C^{14}O_2$. Further, let $V$ represent the amount of $C^{12}O_2$ plus $C^{14}O_2$ converted to acetic acid at any time; $V$ also equals the amount of $C^{12}O_2$ formed from glucose since there is no net change in carbon dioxide. $Va$ is the constant amount of carbon dioxide present throughout the fermentation and $V_f$ is the total carbon dioxide formed or utilized during the experiment.

When a small quantity ($\Delta V$) of $C^{14}O_2 + C^{12}O_2$ is converted to acetic

acid and an equal amount of $C^{12}O_2$ is formed from glucose, the decrease in $C^{14}O_2$ ($-\Delta x$) is given by the expression

$$-\Delta x = \frac{\Delta V}{Va + \Delta V} \cdot x$$

which means that the $C^{14}O_2$ ($x$) is decreased by a fraction equal to the carbon dioxide used ($\Delta V$) divided by the total carbon dioxide present ($Va + \Delta V$). Dividing by $x$ we get

$$\frac{-\Delta x}{x} = \frac{\Delta V}{Va + \Delta V}.$$

In the limit as $\Delta V$ is decreased

$$\frac{-dx}{x} = \frac{dV}{Va + dV} = \frac{dV}{Va}.$$

Integrating this expression between the limits $x_0$ and $x_f$ for $x$, and 0 and $V_f$ for $V$, we find

$$-\int_{x_0}^{x_f} \frac{dx}{x_1} = \ln \frac{x_0}{x_f} = 1/Va \int_0^{V_f} dV = \frac{V_f}{Va}$$

or

$$V_f = 2.3 \cdot Va \log x_0/x_f.$$

$V_f$ has the same units as $Va$. $V_f$ must be divided by the quantity of glucose fermented (= moles of acetic acid formed divided by 2.65) to give carbon dioxide production per unit of glucose.

When this method of calculation is applied to the data of experiment 1 (table 1) where $x_0 = 117$, $x_f = 5.7$, $Va = 0.113$ mM and the glucose decomposed is 0.3 mM, a result of 1.14 moles of carbon dioxide per mole of glucose fermented is obtained. This figure almost certainly errs on the low side because of the likelihood that $x_f$ was raised by contamination with $C^*O_2$ from the gas phase of the culture vessel. Such contamination might easily have been important in this experiment due to the very low activity of the final carbon dioxide and the high activity of the initial carbon dioxide, some of which may have remained in the gas phase or in the Oxsorbent seal during the fermentation. When $x_f$ is very small in relation to $x_0$, a relatively slight increase in $x_f$ will cause a disproportionately large decrease in the factor $\log x_0/x_f$ and therefore in $V_f$.

To obtain a more reliable value for carbon dioxide production two additional experiments were performed in which the ratio $x_0/x_f$ was kept small by limiting the amount of glucose fermented. In this way errors due to $C^{14}$ contamination were made insignificant. From the data of these experi-

ments (table 4) the intermediate carbon dioxide production was calculated to be 2.32 (experiment 2) and 2.19 (experiment 3) moles per mole of glucose fermented.

TABLE 4

THE FERMENTATION OF GLUCOSE IN THE PRESENCE OF $C^*O_2$

Experiments 2 and 3

| | EXPT. 2 | | EXPT. 3 | |
|---|---|---|---|---|
| COMPOUND | MM/10 ML. | CTS./MIN./MM $\times 10^{-3}$ | MM/10 ML. | CTS./MIN./MM $\times 10^{-3}$ |
| Carbon dioxide, initial | (0.191) | 28.5 ± 0.5 | (0.136) | 25.5 ± 0.4 |
| Carbon dioxide, final | 0.191 | 10.0 ± 0.3 | 0.136 | 6.9 ± 0.3 |
| Acetic acid, final | 0.258 | 9.5 ± 0.2 | 0.245 | 8.6 ± 0.2 |
| Glucose fermented | 0.086 | ..... | 0.081 | ..... |

The above results definitely exclude a mechanism such as that implied in reactions (3) and (4). The agreement with the value of 2 to be expected on the basis of reactions (1) and (2) is very satisfactory in the light of the possibility that reactions other than those postulated may be contributing to the dilution of the $C^*O_2$. It should be noted in this connection that only about 85% of the glucose is accounted for as acetic acid; the remainder goes into cell material and unidentified non-volatile compounds.

It may be concluded that the "acetic acid fermentation" of glucose by *Cl. thermoaceticum* involves a partial oxidation of the substrate to two moles each of acetic acid and carbon dioxide followed by a reduction and condensation of the carbon dioxide to a third mole of acetic acid. *Cl. thermoaceticum* is the third species of Clostridium that has been shown to use carbon dioxide in this way.

The authors are greatly indebted to Prof. E. O. Lawrence and members of the staff of the Radiation Laboratory and to Dr. T. H. Norris of the Department of Chemistry for facilities used in the preparation of $C^{14}$.

* Present address: Mallinckrodt Institute of Radiology, Washington University, St. Louis, Mo.

† The authors are indebted to Dr. T. H. Norris of the Dept. of Chemistry of the University of California for the data on self-absorption.

[1] Fontaine, F. E., Peterson, W. H., McCoy, E., Johnson, M. J., and Ritter, G. J., *Jour. Bact.*, **43**, 701–715 (1942).

[2] Barker, H. A., these PROCEEDINGS, **30**, 88–90 (1944).

[3] Wieringa, K. T., *Leeuwenhoek*, **3**, 1–11 (1936).

[4] Wieringa, K. T., *Ibid.*, **6**, 251–262 (1939–1940).

[5] Barker, H. A., Ruben, S., and Beck, J. V., these PROCEEDINGS, **26**, 477–482 (1940).

[6] Ruben, S., and Kamen, M. D., *Phys. Rev.*, **57**, 549 (1940).

[7] Kamen, M. D., unpublished data.

[8] Slade, H. D., Wood, H. G., Nier, A. O., Hemingway, A., and Werkman, C. H., *Jour. Biol. Chem.*, **143**, 133–145 (1942).

# THE CRYSTALLINE PHASES OF SOAP

BY M. J. BUERGER, L. B. SMITH, F. V. RYER AND J. E. SPIKE, JR.
MINERALOGICAL LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, AND LEVER BROS. RESEARCH LABORATORY

Communicated June 14, 1945

*Introduction.*—The view has recently been expressed that soap exists in only four solid phases,[6] known as $\alpha$, $\beta$, $\delta$ and $\omega$, and that only three of these, namely $\beta$, $\delta$ and $\omega$, are encountered in commercial soaps. This simplified scheme does not agree with the results of our own studies of the crystalline phases of soap. We believe that at least part of the error of this view is due to the fact that some of the supposedly simple phases described are actually composites, and that part of it is due to limited knowledge of soap phases themselves. Since we feel that the understanding of soap phases is in a very confused state, we take this opportunity to present information which will tend to clear the existing confusion.

*Orthodox Methods of Producing Various Phases.*—Thiessen and Stauff[1] were the first to point out that sodium stearate and sodium palmitate occur in more than one form. They obtained a form, which they designated $\alpha$, by crystallization from an alcohol solution. They showed that if this form of sodium stearate is heated, it undergoes an irreversible transformation at 52°C. to another form which they called $\beta$. De Bretteville and McBain[2] produced a third form, which they called $\gamma$, to conform with the notation of Thiessen and Stauff, by precipitating sodium stearate from 95% alcohol and drying at 105°C. Immediately after this we[3] showed that the forms known as $\alpha$ and $\beta$ were actually different hydrates, and that the $\alpha$ form can be transformed into $\beta$, $\gamma$, or a still different form (not then dignified by a label) by heating it within the temperature brackets[4] 52–105°, 105–117° and above 117°C., under otherwise atmospheric conditions. For sake of a label we now propose the designation $\sigma$ for the form assumed by heating neutral sodium stearate above 117°C.

Shortly after this, Ferguson, Rosevear and Stillman[5] recognized the existence of only the four phases $\alpha$, $\beta$, $\omega$ and $\delta$. They obtained their $\omega$ phase by cooling neat soap without agitation. They found this phase to occur in commercial framed soaps, most commercial milled soaps, and to predominate in cocoanut oil soaps. They also found it to be the only stable phase in sodium oleate soaps. Another phase which they believed to be identical with Thiessen and Stauff's $\beta$, and which they therefore also called $\beta$, was formed in those soaps which would otherwise crystallize to form $\omega$, but which were agitated at certain temperatures where it was found that $\beta$ was stable. The fourth form, $\delta$, they obtained by several methods, including quenching a hydrous melt, mixing $\alpha$, $\beta$ or $\omega$ with water at room temperature, or extrusion of hydrous sodium palmitate from an orifice at room temperatures.

*New Methods of Producing Different Soap Phases.*—Soap makers had realized for some time that soaps differed in properties according to water content, and, more recently, Bodman[7] discovered that they differed in properties according to temperature of treatment as well as water content. We conceived that the differences so manifested in commercial soaps were to be accounted for by different content of crystalline phases with different conditions of preparation. At a very early date we proved that this was indeed the case by finding that the diffraction patterns of the several commercial varieties of soap were different and that the diffraction effects arose from crystalline phases. This immediately suggested that the phase diagrams of soap-water systems governed the phase mixtures found in commercial soaps, and so influenced the properties of soaps. Our subsequent experimental work proved this to be true. We therefore explored a number of soap-water systems for phase distribution. Before entering into this, we wish to discuss the concepts of *descendent phases* and *phase maps*.

*Descendent Phases.*—At this point we wish to emphasize the distinction between a phase stable in a region where it is developed and the phase or phases which descend from it as a result of changing conditions, such as lowering the temperature in order that the phase can be examined under normal "room temperature" conditions. The former phase properly belongs on a phase diagram. The latter phases, here designated *descendent phases*, are not necessarily the same phase which formed under the original conditions, but are the phases which form from the original one during the change in environment. If no transformations occur during the change, then the original phase and the descendent phase are identical; if a transformation does occur, then they are different. As an example of this relation, McBain's subwaxy phase[4] is the original phase, the $\sigma$ phase is its descendent.

*Phase Maps.*—If the phases of a phase diagram are replaced at all points by their descendent phases, we designate the resulting diagram as a *phase map*. In other words, a phase map is a diagram showing the room temperature equivalents of the phases on the phase diagram.

Clearly, a phase map of a soap-water system is a type of representation of the phase diagram of the system. It should reveal possible new room temperature soap phases at the same time giving some idea of their relationships to one another. To study this, we explored the phase maps of the following individual pure soaps with water:

| SOAP | COMPOSITION | SYMBOL |
|---|---|---|
| Sodium stearate | $NaC_{18}H_{35}O_2$ | $NaC_{18}$ |
| Sodium palmitate | $NaC_{16}H_{31}O_2$ | $NaC_{16}$ |
| Sodium myristate | $NaC_{14}H_{27}O_2$ | $NaC_{14}$ |
| Sodium laurate | $NaC_{12}H_{23}O_2$ | $NaC_{12}$ |
| Sodium caprate | $NaC_{10}H_{19}O_2$ | $NaC_{10}$ |
| Sodium oleate | $NaC_{18}H_{33}O_2$ | $NaC_{\overline{18}}$ |

Equilibrium was attained by the process of severely working the sample. This imposes intense plastic deformation on any solid phases. We have found that this provides the necessary activation energy in mechanical form to induce transformations which may otherwise be inhibited.

*New Phases.*—In this study we encountered seven distinct crystalline soap phases (in addition to several other phases closely related to these). These seven are distributed among the several soap-water systems according to the scheme in table 1.

TABLE 1

THE DISTRIBUTION OF PHASES IN THE PHASE MAPS OF NEUTRAL SODIUM SOAPS WITH WATER

| | $NaC_{10}$ | $NaC_{12}$ | $NaC_{14}$ | $NaC_{16}$ | $NaC_{18}$ | $NaC_{18}^{=}$ |
|---|---|---|---|---|---|---|
| $\kappa$ | x | x | x | .. | .. | .. |
| $\zeta$ | .. | x | x | x | x | .. |
| $\mu$ | .. | .. | x | .. | .. | .. |
| $\epsilon$ | .. | .. | .. | x | x | .. |
| $\delta$ | .. | . | .. | x | x | .. |
| $\alpha$ | .. | .. | .. | .. | x | .. |
| $\eta$ | .. | .. | .. | .. | .. | x |

It will be noted that Ferguson, Rosevear and Stillman's[5] $\omega$ and $\beta$ phases are missing from this list. The significance of this is that each of these phases is composite, and we resolve them into the following components:

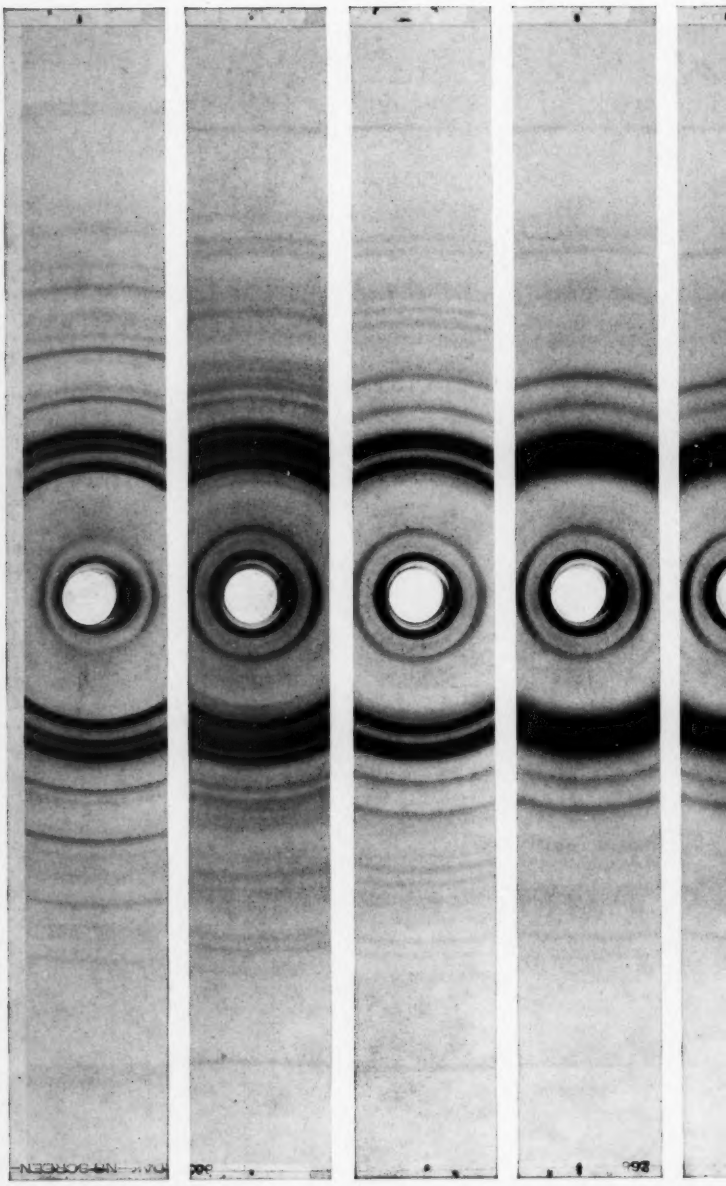
| FERGUSON, ROSEVEAR AND STILLMAN'S PHASES | | OUR PHASES |
|---|---|---|
| $\omega$ | → | $\kappa$ and $\eta$ |
| $\beta$ | → | $\epsilon$ and $\zeta$ |

Ferguson, Rosevear and Stillman's descriptions of conditions under which their phases occur include conditions under which both components appear. Furthermore, Ferguson's "characterizing diffraction ring" does not distinguish between the components in the case of either of their phases $\omega$ or $\beta$. That the components are indeed individual phases is attested by the fact that they occur individually in certain fields of certain phase diagrams and together in others. We accordingly reject Ferguson, Rosevear and Stillman's $\omega$ and $\beta$ phases as being each resolvable into at least two phases.

EXPLANATION OF FIGURE 1

X-ray powder photographs of sodium soap phases. These photographs, which are reproduced here to exactly natural size, were made with a camera having the standard diameter of 114.6 mm. (on which 1-mm. film distance corresponds to 1° for the deviation angle $2\theta$, and to $^1/_2$° for the Bragg angle, $\theta$).

1. $\alpha$ sodium stearate
2. $\delta$ sodium stearate
3. $\epsilon$ sodium stearate
4. $\zeta$ sodium stearate
5. $\beta$ sodium stearate
6. $\mu$ sodium myristate
7. $\kappa$ sodium myristate
8. $\gamma$ sodium stearate
9. $\sigma$ sodium stearate
10. $\eta$ sodium oleate



$\alpha$ $\delta$ $\epsilon$ $\zeta$

(See op

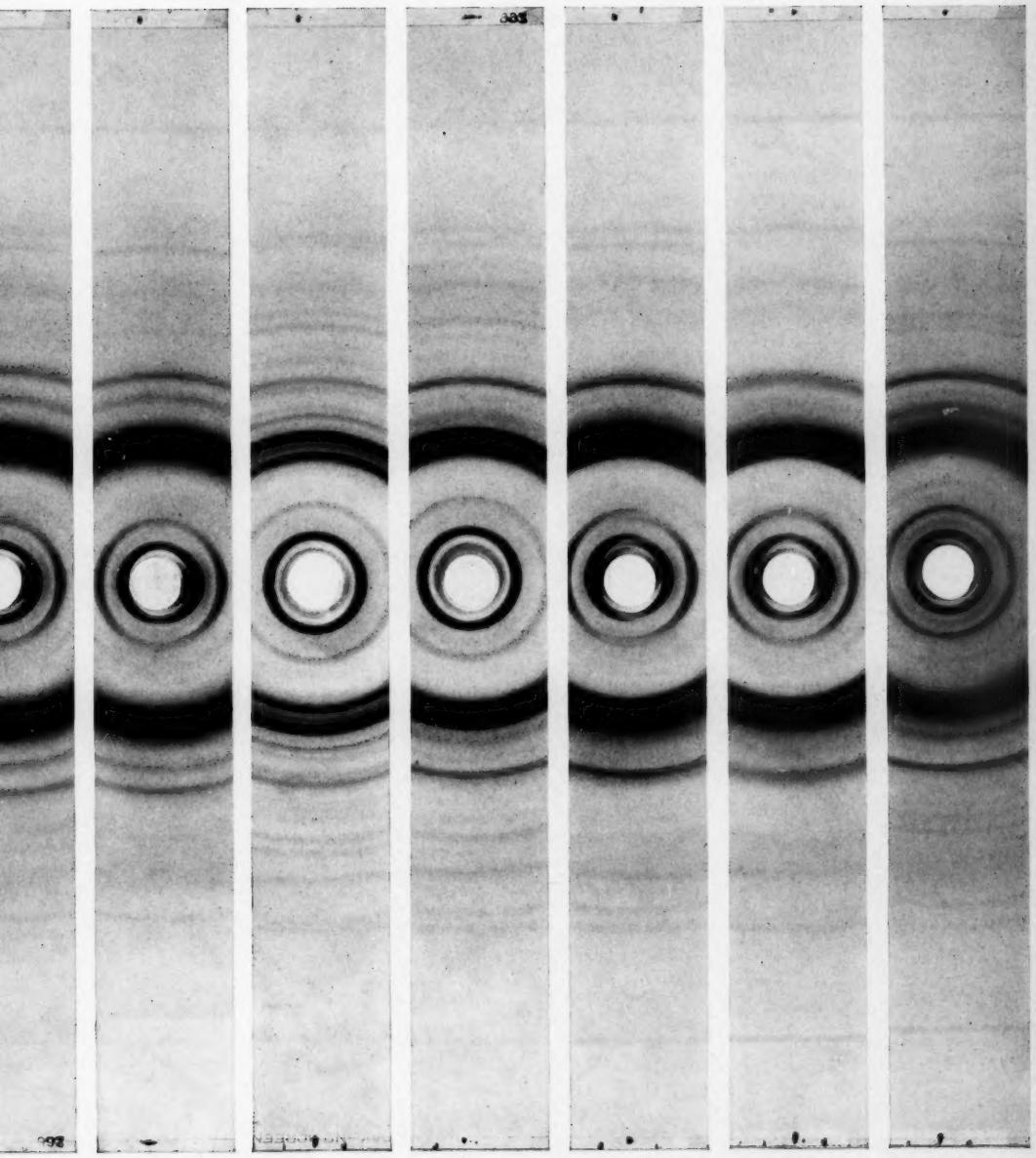



Figure 1

(See opposite page for explanation)

(If the single "characterizing diffraction ring" were the only distinguishing charactertistic of a phase, it would be impossible, with Ferguson's $\omega$ data, to distinguish between the four phases $\kappa$, $\eta$, $\gamma$ and $\sigma$.)

*The Rôle of Hydration.*—Elsewhere[8] it is shown that all the phases of table 1, which occur in the central areas of phase maps, are hydrates. The amount of hydration is a definite fraction of a mole of water per mole of soap. The presence of definite amounts of water in soap crystals which form in a water environment is a consequence of the requirement that the alkali atom of the soap molecule surrounds itself with its normal coördination sphere of oxygen atoms.[9]

There exist two classes of transformations accompanying the dehydration of soap phases. One class (for example, Thiessen and Stauff's $\alpha \rightarrow \beta$) is accompanied by a complete change of crystal structure. The other class is accompanied by comparatively minor alteration in the structure, and consequently is indicated by a comparatively slight change in the x-ray powder pattern of the original phase. This behavior is true of the dehydration of worked $\delta$, $\epsilon$ and $\zeta$.

Since the more and less hydrated modifications of a soap crystal type can be distinguished, even though their powder patterns are rather similar, it is desirable to refer to them in a distinctive manner. We suggest that these forms be distinguished by adding a prime to the symbol of the less hydrated modification. Using this symbolism, we have found that the first step in the dehydration of $\delta$ is $\delta'$, of $\epsilon$ is $\epsilon'$ and of $\zeta$ is $\zeta'$. Thiessen and Stauff's $\beta$ is equivalent to $\zeta'$.

We now call attention to certain relations between Ferguson, Rosevear and Stillman's forms and other forms. They originally* believed[5] their $\beta$ to be identical with Thiessen and Stauff's $\beta$. One component of their $\beta$ is $\zeta$. This component, on dehydration, first loses $^1/_8$ molecule of water and becomes $\zeta'$, which is Thiessen and Stauff's $\beta$. Even if Ferguson's $\beta$ were the pure component $\zeta$, it would not be identical with Thiessen and Stauff's original $\beta$, because the latter is a product of partial dehydration of the former, and the water is lost in a sharp step, giving no evidence of continuity between the phases. The two phases are also readily distinguished by powder pattern.

Ferguson, Rosevear and Stillman[5] express belief that their $\omega$ is identical with de Bretteville and McBain's[2] $\gamma$. This can hardly be the case, even if $\omega$ were a single phase and not resolvable into $\kappa$ and $\eta$, for both of the latter phases are hydrated,[8] while the former is almost, if not quite, anhydrous.

In order that others may identify the several phases involved in soaps, we present in figure 1 a series of representative x-ray powder photographs of the several phase types encountered in sodium soaps. The patterns of the same phase type but different chain length resemble one another basically because the Fourier transforms of the crystals resemble one another. In

detail, the crystals of similar phase type are nearly identical in projection normal to the direction of the chains in the structure. This direction is undoubtedly nearly normal to the sheet structure of soap crystals. Since the plane of the sheets is customarily designated[9] (001), this means that the crystals are nearly similar when projected along zone [001], whose reflections are *hk0*. Thus the (001)* levels of the reciprocal lattices of crystals of different chain length homologues of the same phase are nearly identical and consequently these phases give identical *hk0* contributions to a powder photograph. The reflections involving the *l* index, however (both 00*l* and *hkl*), differ for patterns of the various chain lengths.

*Distribution of Phases on Phase Maps.*—In subsequent publications we plan to describe in detail the phase maps of the several systems of pure soaps with water. At this time we merely outline their several characteristics. In figure 2 is shown the phase map of the system $NaC_{16}$-water, omitting the liquid solution phase which occurs increasingly toward the right of the diagram.

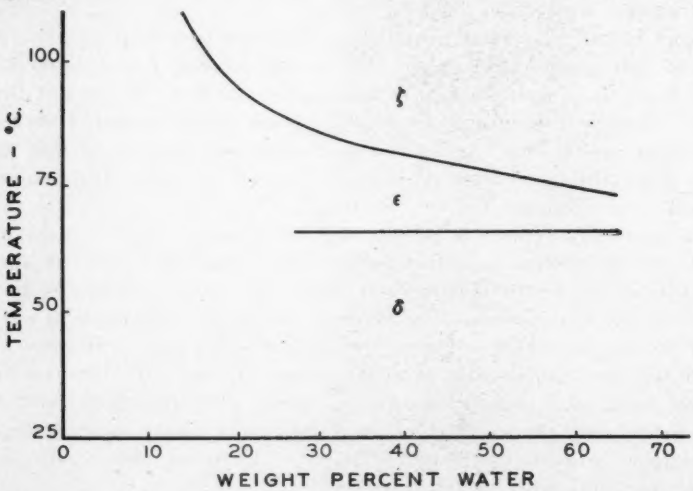



Figure 2

The $NaC_{16}$-water phase map is one of intermediate complexity, but it illustrates some common features of phase maps of simple soap systems. A major feature of such maps is a set of three regions, the middle region being bounded above by a sloping line and below by an almost horizontal line. These boundaries converge toward higher water contents. The upper line, at least near the middle region, checks well with McBain's[10] $T_c$, the temperature of the final melting of the crystalline phases. The lower horizontal phase boundary is a new one and evidently represents a phase

transformation in the crystalline state. In the $NaC_{16}$-water phase map, the $\delta$ phase occupies the field below this boundary, the $\epsilon$ phase occupies the middle field and the $\zeta$ phase occupies the field above the upper boundary. It is visually evident whether a sample has been prepared above or below the $\epsilon$-$\zeta$ boundary. Samples worked in the $\zeta$ field are brittle and have a compact texture, while those worked in the $\epsilon$ field are tougher and have a fibrous appearance with silky luster. The $\delta$-$\epsilon$ boundary corresponds well with the temperature of ready solubility, as listed by McBain and Lee.[10]

Phase maps of water with soaps of longer chain length are somewhat more complex. The $NaC_{18}$-water diagram is similar to the $NaC_{16}$-water diagram, with the following additional complexities. The $\delta$ phase field is a thick, almost horizontal stratum below whose lower boundary occurs another field containing the $\alpha$ phase. Furthermore, the upper field which contains only pure $\zeta$ for $NaC_{18}$ is subdivided for $NaC_{16}$: pure $\zeta$ occurs at low moistures, but $\zeta + \delta$ appears at intermediate moistures. Since soap solution is also present, this appears to be a field of three phases, which cannot coexist in the phase diagram of a two-component system. Evidently $\delta$ is generated from the phases stable in this region when the sample is cooled for examination.

Phase maps of water with soaps of shorter chain length tend to be simpler. In the $NaC_{14}$-water phase map, the upper, middle and lower regions contain $\kappa$, $\zeta$ and $\mu$, respectively. The latter phase occurs only, so far as we know, in this system. (At very low water contents this phase appears to be displaced by $\kappa$.) The middle region tapers off at higher moisture contents toward about 70% water. In the $NaC_{12}$-water map, fewer phases occur. Both upper and lower region are occupied by the $\kappa$ phase; the middle is occupied by $\zeta$ and vanishes at about 75% water. In the $NaC_{10}$-water system, only the $\kappa$ phase occurs, yet the middle region is marked out by slightly different detail in the $\kappa$ powder pattern, as well as by orientation differences.

The phase map of NaOl-water is also very simple, consisting chiefly of the phase $\eta$. The only detail in the map corresponds with a minor pattern change below 6% water. We show elsewhere[8] that the samples having water contents greater than this are actually crystals of the hemihydrate, $NaC_{18}^{=} \cdot {}^1/_2H_2O$. Below 6% water the dehydrated phase $\eta'$ occurs.

*Phase Maps of Commercial Soaps.*—Commercial soaps have complex phase maps which differ from soap to soap. This might be expected, since commercial soaps are accidental and convenient mixtures made from fats containing varying spectra of chain lengths, and containing both saturated and unsaturated components.

We are not in agreement with the statement[6] that commercial soaps contain only the three phases $\beta$, $\omega$ and $\delta$. We find that phase maps of commercial soaps are so complex that we cannot deduce them with any cer-

tainty from powder photographs. In figure 3 we present a phase map of a certain commercial soap. In this map, where we have felt uncertain of identification of the phases of a phase field, we have attempted to indicate the minimum number of pure phase patterns into which we believe the powder photographs can be resolved. We present this diagram with some hesitation, since we are certain we would revise it in light of subsequent knowledge, yet we feel that it should be presented in view of the published statements as to the simplicity of the phase aggregates in commercial soaps.

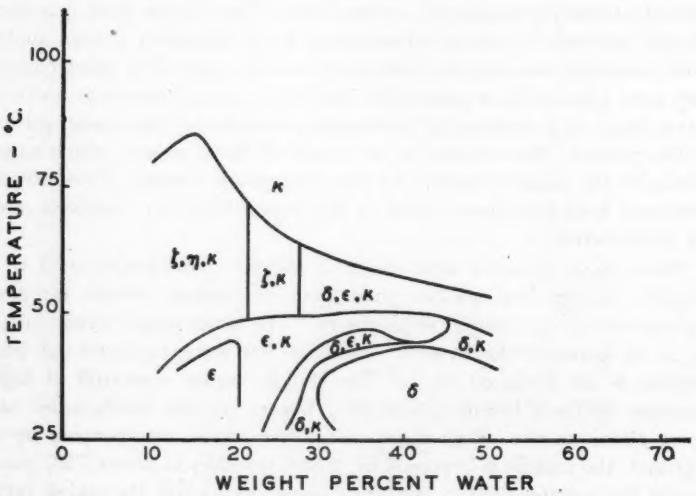



Figure 3

* Subsequently Ferguson[6] expressed the view that his $\beta$ differed from Thiessen and Stauff's $\beta$ by being indefinitely hydrated.

[1] Thiessen, Peter A., and Stauff, Joachim, "Feinbau und Umwandlungen kristallisierter Alkalisalze langkettiger Fettsäuren," *Zeit. Physikal. Chem.* (*A*), **176**, 397–429 (1936).

[2] Bretteville, Alexander de, Jr., and McBain, J. W., "X-ray Evidence for a Third Polymorphic Form of Sodium Stearate," *Science*, **96**, 470–471 (1942).

[3] Buerger, M. J., Smith, L. B., de Bretteville, Jr., A., and Ryer, F. V., "The Lower Hydrates of Soap," *Proc. Nat. Acad. Sci.*, **28**, 526–529 (1942).

[4] McBain, James W., Vold, Robert D., and Frick, Mary, "A Phase Rule Study of the System Sodium Stearate-Water," *Jour. Phys. Chem.*, **44**, 1013–1024 (1940).

[5] Ferguson, R. H., Rosevear, F. B., and Stillman, R. C., "Solid Soap Phases," *Ind. Eng. Chem.*, **35**, 1005–1012 (1943).

[6] Ferguson, R. H., "The Four Known Crystalline Forms of Soap," *Oil and Soap*, **21**, 6–9 (1944).

[7] Bodman, John W., U. S. Patent No. 2,215,539, Sept. 24, 1940.

[8] Gardiner, K. W., Buerger, M. J., and Smith, L. B., "The Hydrate Nature of Soap," *Jour. Phys. Chem.* (in press).

[9] Buerger, M. J., "Soap Crystals," *American Mineralogist* (in press).

[10] McBain, James W., and Lee, W. W., "Vapor Pressure Data and Phase Diagrams for Some Concentrated Soap-Water System Above Room Temperature," *Oil and Soap*, 20, 17–25 (1943).

---

## *WILLARD GIBBS ON SOARING FLIGHT*

BY EDWIN B. WILSON

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated July 12, 1945

In these days when so much depends on aerial supremacy it may interest some to see a hitherto unpublished letter of Gibbs to Langley dating from the very early infancy of aeronautical research in this country. The letter was furnished to me by Professor Ralph G. VanName, nephew of Gibbs.

On April 13, 1894, Langley wrote Gibbs asking for help in understanding soaring flight. He gave the usual Newton-type of formula for the wind pressure upon the aerofoil, proportional to the area, the square of the velocity and a function of the angle of attack which could be taken as proportional to that angle when it was small. Six weeks later Gibbs replied as follows:

New Haven, May 30/94

My dear Professor Langley

I do not know whether the following results—certainly very meager after so long a delay—will be at all to your purpose, but the discussion of a simple case may throw some light on the general question.

Let the velocity of the wind be

$$V = A + a \sin nt. \tag{1}$$

Let the horizontal velocity of the aeroplane (measured *against* the wind) be regulated by varying the inclination so as to be expressed by a similar function, say,

$$v = B + b \cos nt. \tag{2}$$

The *relative* horizontal velocity will be

$$w = C + a \sin nt + b \cos nt, \tag{3}$$

where $C = A + B$. If $\alpha$ is the tangent of the elevation of the edge (supposed small),

$$\alpha = -\frac{dv}{gdt} - \frac{Ew^2}{g}, \tag{4}$$

where $Ew^2$ is the resistance of the air to the edge of the plane divided by its mass; and if $z$ is the vertical height,

$$\frac{dz}{dt} = \alpha w - \frac{g}{Pw} = -\frac{w}{g}\frac{dv}{dt} - \frac{Ew^3}{g} - \frac{g}{Pw}, \tag{5}$$

where $g/Pw$ is the rate at which the aeroplane would settle down when $\alpha = 0$. We shall treat $P$ as roughly a constant. It is here assumed not only that $\alpha$ is small, but also that $w$ is large, and $d^2z/dt^2$ is small compared with $g$, so that the variations of vertical momentum may be neglected in (4). Since by (2)

$$\frac{dv}{dt} = -nb \sin nt, \tag{6}$$

(5), (6) and (3) give

$$\frac{dz}{dt} = \frac{bn}{g} \sin nt\,(C + a \sin nt + b \cos nt) - \frac{Ew^3}{g} - \frac{g}{Pw}, \tag{7}$$

$$\int_{t=0}^{t=2\pi/n} dz = \frac{\pi ab}{g} - \frac{2\pi}{n}\left[\frac{Ew^3}{g} + \frac{g}{Pw}\right] \text{average value.} \tag{8}$$

Continued soaring is possible, if

$$\frac{nab}{2g} > \left(\frac{Ew^3}{g} + \frac{g}{Pw}\right) \text{ave.} \tag{9}$$

Since $C$ is the average value of $w$, we shall err on the side of safety if we write for (9),

$$\frac{nab}{6g} > \frac{EC^3}{g} + \frac{g}{PC}, \tag{10}$$

*provided* that $C$ is three times as great as $a$ and as $b$, and *provided* also that $\alpha$ and $\frac{1}{g}\frac{d^2z}{dt^2}$ are sufficiently small to make our formulae reasonably accurate.

To make $\alpha$ small, $nb$ must be small compared with $g$. To make $d^2z/dt^2$ small compared with $g$, we must have $n^2bC$ small compared with $g^2$. To fix our ideas, we may make $b = C/3$. Then (10) becomes

$$\frac{na}{18g} > \frac{EC^2}{g} + \frac{g}{PC^2}. \tag{11}$$

The left-hand member of this condition is given by the wind. If $E$ is small enough and $P$ large enough, it is easy to satisfy (11) by values of $C$ which

are small compared to $g/n$, which will make soaring possible, and if such values of $C$ or any of them are greater than $A$ (the average velocity of the wind), continued soaring *against* the wind is possible.

The letter is taken from the handwritten copy which Gibbs kept and may not be identical with that which he sent; it closes without the customary "Yours truly" or equivalent and without signature, though undoubtedly both appeared on the letter sent to Langley.

Under date of June 9, 1894, Langley wrote: "I beg to return my very best thanks for your highly valued communication of the 31st ultimo. Not being certain when I should hear from you, I had asked the independent assistance of one or two other gentlemen, whose analytical skill I thought might be more trusted than my own.

"I am chiefly surprised at the entirely different ways in which the problem can be looked at and treated, and yours is certainly original and distinctive. It has arrived too late, I regret, to be used in a communication which is just going to France, but I shall probably make use of it later, with due public acknowledgment of your valued aid and great kindness."

About six months later (January 25, 1895), Langley wrote again:

"You will be interested in the discussion of a mathematical problem in aerodynamics, prepared at my request by Mr. de Saussure independently of the discussion that you were good enough to send me, which I have preserved, and may ask your permission to make use of it at some future time. Mr. de Saussure's article appeared as a supplement to the French translation of my paper on The Internal Work of the Wind in the *Revue de l'Aeronautique*.

"I have sent to you, under a separate cover, a copy of the paper in question, and, should your leisure permit, I shall be glad to hear any criticism upon it that you may make.

"Renewing my thanks for your assistance, I am,"

So far as I have been able to ascertain Langley never took occasion to use Gibbs's material, nor Gibbs to accept the invitation to comment on de Saussure's article—there is certainly no evidence of either in the meager papers Gibbs left at his death.

# EFFECTS OF EXPOSURE TO ULTRA-VIOLET LIGHT ON VISUAL THRESHOLDS*

BY ERNST WOLF

BIOLOGICAL LABORATORIES, HARVARD UNIVERSITY

Communicated July 5, 1945

Effects of ultra-violet light upon the eye are well known in the form of "snow blindness" and in many other pathological appearances found particularly among industrial workers who are exposed to light sources emitting great quantities of ultra-violet.[1, 2, 3] Since the wave band of light is continuous from abiotic rays to the visible part of the spectrum, it is of interest to know the limits of detrimental effectiveness of the short wave-lengths.

The transmission of the ocular media has been studied by scores of investigators and the observations vary considerably. We may assume, however, that some human and other vertebrate eyes transmit wave-lengths as low as 310 mμ. Recent visual tests in the ultra-violet[4] indicate that light of 302 mμ is perceived by very young subjects (6–10 years). With progressing age the low limit moves to longer wave-lengths, middle-aged persons rising to 310–320 mμ and at higher age to 360 mμ and higher. In laboratory tests a decrease in visual acuity after exposure to ultra-violet has been found in monkeys and man.[5] No conclusions concerning wave-length limits causing these effects can be drawn from these data. For industry, suggestions regarding safety measures have been made in accordance with the assumption that only wave-lengths shorter than 305 mμ would be detrimental to the visual mechanism and that crown or flint glass of sufficient thickness would provide adequate protection.[6] A systematic study of the effects of different wave bands below 400 mμ on the vision of some suitable animal material which would permit comparison with similar effects upon the human eye seems therefore desirable.

Responses to flicker have been studied in a great variety of vertebrates and flicker functions established.[7] In animals the end-point of flicker recognition at a given intensity of light is determined by the appearance of a head nystagmus due to the perception of a cylindrical system of alternate translucent and opaque stripes rotating around the animal.[8] While the test method differs from that customarily used for man, the flicker function obtained in this fashion is identical with such functions determined by other methods.[9, 10] The behavior of flicker thresholds in relation to light intensity permits therefore in any suitable organism an investigation of visual function under varied experimental conditions, for instance after exposure to ultra-violet light. If by the ultra-violet radiation any functional changes are induced, it is possible to demonstrate (*a*) any deviation

of a response threshold from its normal level at any given point along the flicker function, (*b*) its magnitude of deviation in relation to exposure time and spectral region employed and (*c*) the course of return of the threshold to its normal level with time.

For test, an animal was chosen which would fulfil the following conditions: (*a*) availability in quantity at all seasons, (*b*) genetic uniformity, (*c*) precision and reproducibility of a threshold response and (*d*) an organism that would without aid keep its eyes open during exposure to ultra-violet, so that a full effect of irradiation is assured. For the purpose baby chicks, 3–15 days old, of white Leghorn stock, proved particularly suitable.

The eye of the chick transmits at least as far as 315 m$\mu$. The visual purple of the chick and its absorption spectrum are identical with those of man.[11] The chick's eye possesses, however, embedded in the retina considerable numbers of colored oil droplets which may act as additional light filters and which are absent in the human eye.[12] But since such a screen would rather reduce than enhance the action of ultra-violet, it is justifiable to assume that in other eyes the ultra-violet effect might be more pronounced.

For normal chicks a flicker function has been established[13] which bears the typical characteristics of such curves for other visually duplex vertebrates.[14] The critical flicker frequency is a function of the intensity of illumination, increasing in a double S-shaped form with light intensity, according to the participation of rods and cones in the visual process (Fig. 1). For testing F = 30/sec. with a light/dark ratio of 1:1 in each flicker cycle has been chosen, and the change of threshold value studied after the chick's exposure to ultra-violet. The testing point lies in the midregion of the curve and at an intensity level where responses are precise and easily recognizable by the observer.

While the chicks (4 to 6 at one time) are confined to a cylindrical wire mesh cage, but free to move, with access to food and water, they are exposed to the intense radiation of three 250-watt mercury vapor lamps of the GE type H-5 which are mounted at eye level equally spaced outside the cage and surrounded by a reflecting surface. Lamp housings permit the insertion of filters for selection of specific spectral regions. For exposure, a standard period of 60 minutes was chosen.

Each individual chick is then transferred to a glass cage to be tested at F = 30/sec. for the onset of the nystagmus response in dependence on light intensity (white light), and the tests are repeated at intervals until the threshold returns to its normal level. The first reading is taken 60 minutes after cessation of exposure to allow for complete dark adaptation. Dark adaptation should be complete after 40 to 45 minutes;[15] any deviation of the threshold from its normal level should therefore be due to the exposure to the ultra-violet and not to incomplete adaptation. When

exposing for control some chicks, not previously exposed to ultra-violet, to *white* light (Tungsten filament lamps), equal in intensity to the three mercury vapor lamps, and after proper dark adaptation, no effects on the threshold can be noticed. After ultra-violet treatment the threshold is, however, much higher, depending upon the extent of the spectral region employed.

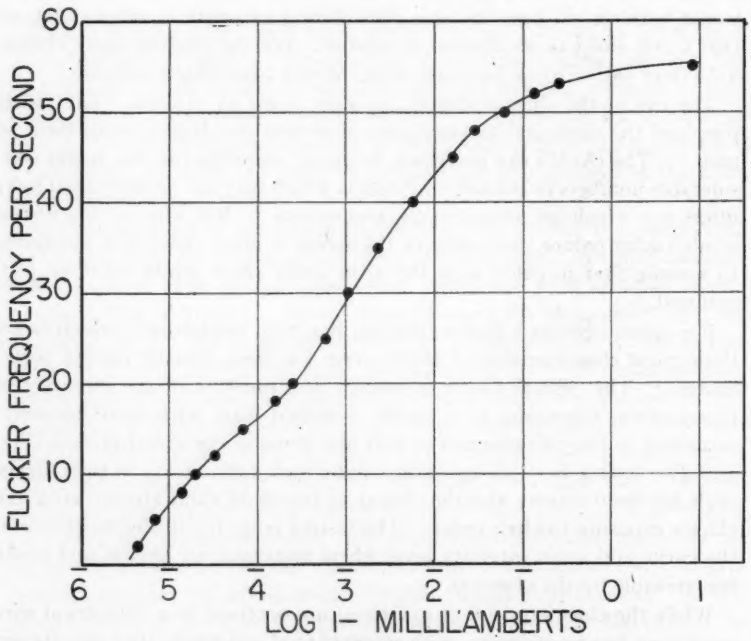



FIGURE 1

Relations between flicker frequency ($F$) and mean critical intensity of light ($I$) for response to flicker in the chick.

Chicks receiving the full impact of the mercury vapor lamps from which the surrounding glass jackets have been removed may show conjunctivitis, keratitis and hyperemia of corneal capillaries, while the fundus looks normal. These pathological changes are, however, absent as soon as the quartz mercury vapor tubes are shielded (as they usually are) by their envelopes of Corning glass No. 774. A study of the behavior of the flicker recognition end-point reveals, however, that it is greatly altered and takes many hours to return to its normal level. It is, therefore, evident that the exterior appearance of the eye does not suffice to decide whether the visual mechanism has been affected by the ultra-violet.

The spectral range of particular interest lies between the lowest limit of transmission of ultra-violet by the eye (>300 mμ) and the beginning of the visible range (<400 mμ). By the choice of suitable filters this region is subdivided to ascertain the specific ultra-violet bands which affect the visual threshold. The filters according to their order of transmission are: Corning 774, AO crown glass 1045, ordinary plate glass, AO Cruxite 1794, Corning 3850, AO Calobar C 1827, AO 2614, Polaroid XY 91 VO and Corning 3389. Their transmission curves, determined with a Beckman spectrophotometer, are given in figure 2.

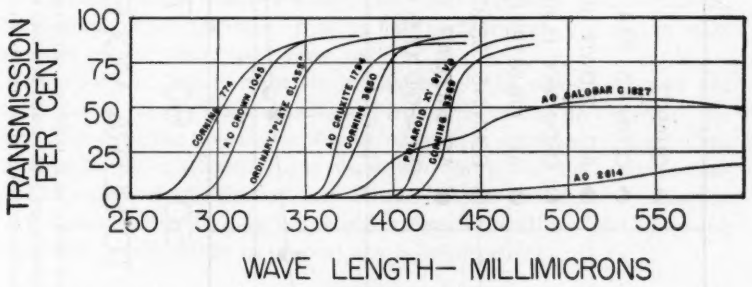



FIGURE 2

Spectrophotometric transmissions of filter glasses used for elimination of the ultra-violet part of the spectrum.

With the mercury vapor as light source and shielded by quartz only, the first threshold recorded one hour after cessation of exposure is about 45 times higher than the normal. Repeated tests show a gradual lowering of the threshold with time, until the normal level is reached not before 72 hours (3 days). In this case pathological changes are clearly visible and some damaging effects were to be expected. With Corning 774 as a filter, a smaller initial threshold change is obtained. This filter begins to transmit at 260 mμ and thus excludes the strong 254 mμ mercury line. In this case the amount of light necessary for threshold recognition is 5 times the normal, and the approach to the normal sensitivity level is complete after about 30 hours. By gradually reducing the extent of the ultra-violet band from 290 mμ to 360 mμ through the use of crown, plate glass, Cruxite and Corning 3850, the initial threshold change and the subsequent recovery are found to be direct functions of the gradual elimination of the short waves. For crown glass the initial threshold is 4 times, for plate glass 3 times, for Cruxite 1.5 times and for Corning 3850 1.3 times higher than normal; and for crown, plate, Cruxite and Corning 3850 the final level is reached after 24, 18, 9 and 6 hours, respectively.

With filters which do not transmit any appreciable amount of light at 365 mμ and below, the picture changes abruptly. The first threshold

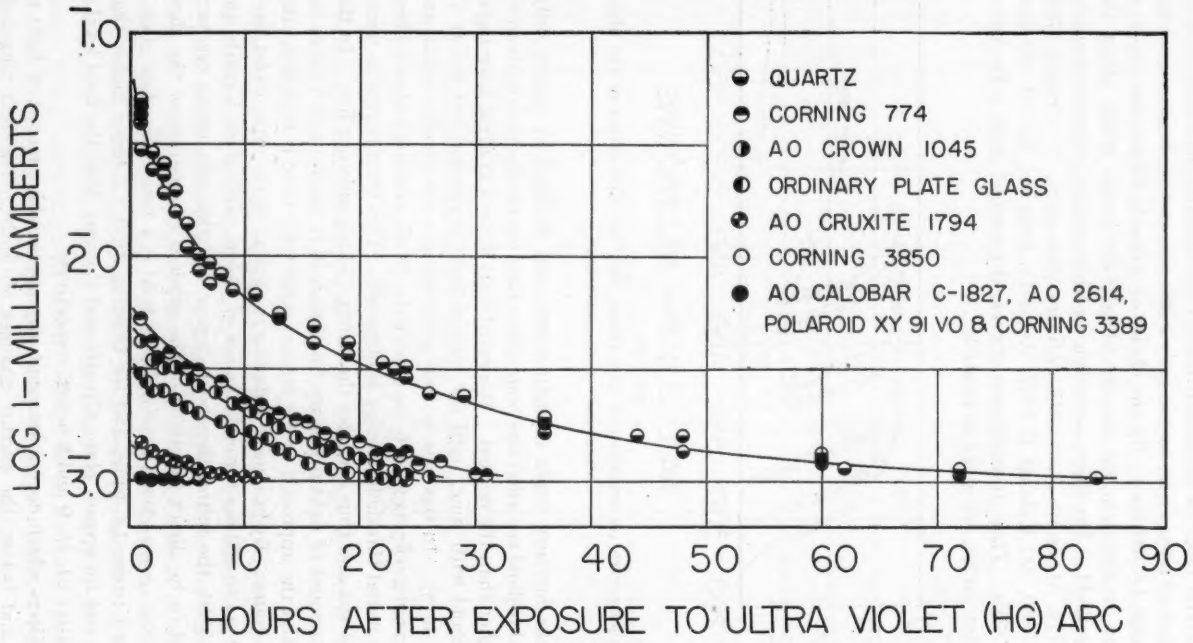



FIGURE 3

**Determinations of threshold intensities of light necessary for recognition of flicker at F = 30/sec. after exposure to ultra-violet light of different spectral regions. One hour after completion of exposure the thresholds are highest and decrease with time until a final level is reached. The initial deviation from the normal and the time for recovery are functions of the extent of the ultra-violet spectrum.**

determination one hour after completion of exposure is at the normal level and does not change in repeated tests. AO 2614 begins to transmit at 380 mμ, Polaroid XY 91 VO and Corning 3389 at 400 mμ and above. Calobar transmits about 0.5% at 360 mμ and 3.1% at 370 mμ. The mercury spectrum has very strong lines at about 365 mμ which probably represent the limit of injurious radiation. For Calobar the relative energy transmitted in this region is only one-twentieth of that transmitted by Corning 3850; consequently a small ultra-violet effect is apparent with the Corning filter and not with Calobar.[16] The data are shown in detail in figure 3. Each point represents the mean of 4 to 6 individual threshold determinations. The deviation from the mean is in no case greater than $^1/_{10}$, and in most cases not greater than $^1/_{20}$ of a log. unit.

*Summary.*—If, based upon the similarity of the transmission of the ocular media, conclusions can be drawn from the findings on the eye of the chick, it is apparent that under absence of exterior pathological conditiöns, recognizable by ophthalmoscopic inspection, the visual mechanism is impaired by ultra-violet light between 300 and 365 mμ. Protective means for the eye should, therefore, be filters which absorb the ultra-violet up to 365 mμ totally or sufficiently to prevent any injurious effects.

* This research was supported by a grant of the American Optical Company.

[1] Laurens, H., *The Physiological Effects of Radiant Energy*, Chemical Catalog Co., New York, 1933.

[2] Blum, H. F., *Photodynamic Action and Diseases Caused by Light*, Reinhold, New York, 1941.

[3] Rieke, F. E., *Jour. Am. Med. Assoc.*, **122**, 734 (1943).

[4] Friedrich, W., and Schreiber, H., *Arch. ges. Physiol.*, **246**, 621, 790 (1943).

[5] Verhoeff, F. H., and Bell, L., "The Pathological Effects of Radiant Energy on the Eye," *Proc. Am. Acad. Sci.*, **51**, No. 13 (1916).

[6] Kinsey, V. E., Cogan, D. G., and Drinker, P., *Jour. Am. Med. Assoc.*, **123**, 403 (1943).

[7] Crozier, W. J., Wolf, E., and Zerrahn-Wolf, G., *Jour. Gen. Physiol.*, **20**, 211 (1936); **21**, 17 (1937). Crozier, W. J., and Wolf, E., *Ibid.*, **21**, 313 (1938); **22**, 311, 436, 555 (1939); **23**, 229, 667 (1940); **24**, 317 (1940), 585 (1941); **25**, 331 (1942); **27**, 315 (1944).

[8] Crozier, W. J., and Wolf, E., *Ibid.*, **23**, 531 (1940).

[9] Crozier, W. J., Wolf, E., and Zerrahn-Wolf, G., *Ibid.*, **21**, 203 (1937).

[10] Crozier, W. J., and Wolf, E., *Ibid.*, **27**, 287 (1944).

[11] Wald, G., and Zussman, H., *Jour. Biol. Chem.*, **122**, 449 (1938).

[12] Franz, V., *Handbuch. d. vergl. Anat. d. Wirbeltiere*, Berlin, 1934. Krause, R., *Mikroskopische Anatomie der Wirbeltiere*, Vol. 2, Berlin, 1922. Walls, G. L., *The Vertebrate Eye*, Cranbrook Inst., Bloomfield Hills, Michigan, 1942. Wood, C. A., *The Fundus Oculi of Birds*, Lakeside Press, Chicago, 1917.

[13] Data to be published in detail elsewhere.

[14] Crozier, W. J., Wolf, E., and Zerrahn-Wolf, G., these PROCEEDINGS, **24**, 125, 438 (1938).

[15] Hecht, S., "The Nature of the Photoreceptor Process," Handbook of Gen. Exp. Psychology, Clark University Press, Worcester, 1934.

[16] Filter glasses transmitting ultra-violet and excluding the visible part of the spectrum produce ultra-violet effects similar to those presented above. The details will be discussed in a later publication.

# ON PLATEAU'S PROBLEM WITH FREE BOUNDARIES

BY R. COURANT

NEW YORK UNIVERSITY

Communicated July 5, 1945

Plateau's problem with free boundaries concerns minimal surfaces $M$ of least area with parts of the boundary prescribed as Jordan arcs while other "free" parts of the boundary are merely restricted to prescribed boundary surfaces $S$. The existence of such minimal surfaces has been established,[1] but the question remains open whether specific statements concerning the "*trace*" of the minimal surface $M$ on the prescribed boundary surfaces $S$ can be made, "trace" meaning the set of boundary points of $M$ on $S$. Examples show that this trace need not be a continuous curve;[2] therefore, to secure a "reasonable" trace of $M$, conditions have to be imposed on $S$.

To fix the ideas we consider the case of a doubly connected minimal surface $M$ of least area with its free boundary on a closed surface $S$ and its fixed boundary on a closed Jordan curve $\Gamma$ outside $S$; $M$, lying in the $\mathbf{x}$-space, may be parametrically represented by an harmonic vector $\mathbf{x}(u, v)$ with components $x, y, z$ where $u, v$ ranges over an annular ring $\bar{M}$ in the $u, v$ plane between the unit circle $\bar{\rho}$ and a concentric circle $\bar{\Gamma}$, so that $\bar{\Gamma}$ is mapped onto $\Gamma$ and that $\bar{\rho}$ corresponds to the trace $\rho$ of $M$ on $S$.

It is the purpose of the present note to show:[3]

(*a*) *The trace is a continuous curve if the prescribed boundary surface $S$ is convex.*

(*b*) *The trace is a rectifiable curve if $S$ is convex and if there exists a cone of supporting planes of $S$ bounding together with $S$ a portion of the space which contains $\Gamma$.*[4]

We first prove the theorem (*a*): By a basic elementary lemma[5] one can draw in $\bar{M}$ circular arcs $\bar{\gamma}$ of arbitrarily small radii with any point $\bar{p}$ on $\bar{\rho}$ as center, joining two points on $\bar{\rho}$, such that the images $\gamma$ in the $\mathbf{x}$-space are arbitrarily short (and analytic, except possibly at the end-points). The arc $\bar{\gamma}$ as well as its image $\gamma$ on $M$ is called a *bridge*.

The small pieces of the minimal surface $M$ corresponding to the small biangular part of $\bar{M}$ between $\bar{\gamma}$ and $\bar{\rho}$ is denoted by $M(\gamma)$. Then we prove by indirect reasoning: For sufficiently small $\epsilon$ there exists a $\delta(\epsilon)$ so that, for all points $\bar{p}$ on $\bar{\rho}$, the piece $M(\gamma)$ is confined in a sphere of radius $\epsilon$ provided that the bridge $\gamma$ is shorter than $\delta(\epsilon)$. If this statement were not true there would exist a sequence of bridges $\bar{\gamma}$ and $\gamma$ for which the lengths tend to zero and for which the diameter of $M(\gamma)$ remains above a positive bound $c$. We may assume that the bridges $\gamma$ converge to a point $O$ on $S$. About $O$ we draw a sphere $\Sigma$ of a small radius $r$ which encloses the small bridge. Then we show—and this is sufficient for the proof of our

statement (*a*)—that $M(\gamma)$ lies entirely in $\Sigma$. To this end we exploit the assumed minimum area property of $M$ by performing a geometrical construction which would provide an admissible surface with area less than that of $M$ if we assume that $M(\gamma)$ has points outside of $\Sigma$.

Making this assumption we replace $M(\gamma)$ by a surface $M_0(\gamma)$ which consists of the part of $M(\gamma)$ inside $\Sigma$ and in addition of that part of $\Sigma$ formed by points onto which points of $M(\gamma)$ outside of $\Sigma$ and visible from $O$ are projected from $O$. By normal projection onto a sphere the area of a surface outside is diminished;[6] and because of the convexity of $S$ the surface $M_0(\gamma)$ joins $\gamma$ with $S$; hence the surface $M - M(\gamma) + M_0(\gamma)$ would be an admissible surface in the original minimum problem with area less than that of $M$, contrary to our assumption. Thus theorem (*a*) is proved.

The proof of theorem (*b*) requires a more elaborate reasoning. It proceeds in the following steps:

1. The given surface $S$ is approximated by a sequence of convex polyhedra $K$ for which again cones of supporting planes exist as specified above. Without restricting the generality of the result we assume that at every vertex of $K$ three faces converge and form three obtuse angles.

2. With such a polyhedron $K$ instead of $S$ as the free boundary the minimal surface $M$ is shown to have on $K$ a trace $\rho$ which consists of a finite number of analytic arcs.

3. The assumption of theorem (*b*) permits the conclusion—and this is the key to the proof—that for a sequence of polyhedra $K$ approaching $S$ the length $L$ of the piecewise analytic trace $\rho$ remains uniformly bounded.

4. Now one may choose a subsequence of polyhedra $K$ approaching $S$ such that the boundaries $\rho = \rho_K$ converge to a rectifiable curve $\rho$ on $S$. According to the established theory, the corresponding minimal surfaces converge to a minimal surface which solves the problem for the free boundary surface $S$ and has as trace on $S$ the rectifiable curve $\rho$.

Incidentally, by Fatou's theorem, the *tangent plane* on $M$ in a point $P$ has a *limiting position* for $P$ approaching almost every point on $\rho$ and this position is *normal* to $S$.

As to the details, the first step, being of a quite elementary nature, is not amplified here. Let us for the moment assume the statement (2), then (3) can be proved as follows: First: The—piecewise analytic—trace $\rho$ of $M$ on $K$ (and the whole surface $M$) lies together with $\Gamma$ inside of the cone formed by the supporting planes. For, otherwise the minimal surface $M$ of least area connecting $\Gamma$ with $K$ would protrude across one of the supporting planes, say the plane $\Pi$. Since $K$ is convex and since $M$ rests on $K$ this situation would imply the existence of a supporting plane to $M$, parallel to $\Pi$, which is incompatible with the fact that $M$ has negative curvature.

If $\theta$ and $r$ denote polar coordinates in the annular ring $M$ then $\mathbf{x}(u, v)$ remains analytic on $\bar{\rho}$, i.e., along $r = 1$, except for a finite number of points. As known from the established theory $M$ is orthogonal to $K$ at points of $\rho$ on a face; likewise along a portion of $\rho$ coinciding with an edge of $K$ the normals to $M$ lie in a supporting plane to $K$ through this edge.[7] If $\bar{\rho}$ is monotonically described the image point on $\rho$ will describe $\rho$ monotonically, as follows from the established theory.

Omitting the finite number of points on $\bar{\rho}$ in which the analyticity of $\mathbf{x}(u, v)$ is interrupted, i.e., which correspond to points on $\rho$ in which an edge or vertex is reached from a face, we now recognize $\mathbf{x}_r$ along $\rho$ as a vector normal to a face or to a supporting plane of $K$. Taking the vertex of our cone of supporting planes as origin in the $\mathbf{x}$-space, the vectors $\mathbf{x}$ and $-\mathbf{x}_r$ form an acute angle not exceeding $\pi/2 - \alpha$, $\alpha$ being a positive quantity. Hence

$$-\mathbf{x}\mathbf{x}_r > |\mathbf{x}|\ |\mathbf{x}_r| \frac{1}{\sin \alpha}.$$

Since along $\rho$ the distance $|\mathbf{x}|$ is bounded from below we have along $\rho$ an inequality of the form

$$|\mathbf{x}_r| < -A\ \mathbf{x}\mathbf{x}_r \qquad (1)$$

with a value of $A$ which can be taken as the same for all minimal surfaces $M$ belonging to polyhedra $K$ sufficiently near to $S$ in our approximation.

As a sequence of polyhedra $K$ tends to $S$ the corresponding annular rings $\bar{M}$ converge to a limit ring. $\bar{M}'$ may denote the ring between $\bar{\rho}$ and the middle circle $\bar{\rho}'$ of the ring; then $\mathbf{x}$ is analytic on $\bar{\rho}'$ and $|\mathbf{x}|$ and $|\mathbf{x}_r|$ are bounded along $\bar{\rho}'$. Since the area of $M = M_K$ converges to the area of $M = M_S$, the established theory of minimal surfaces implies that

$$D(\mathbf{x}) = \int\int_{\bar{M}} (\mathbf{x}_u^2 + \mathbf{x}_v^2)\, du dv < A_1$$

where also $A_1$ may be chosen as the same bound for all polyhedra $K$ under consideration. It follows by a well-known reasoning that not only $|\mathbf{x}|$ but also $|\mathbf{x}_r|$ on $\bar{\rho}'$ is bounded by the same constant $A_2$ for all polyhedra considered.

Using Green's formula and $\Delta\mathbf{x} = 0$ we have for the half ring

$$A_1 > \int\int_{\bar{M}} (\mathbf{x}_u^2 + \mathbf{x}_r^2)\, du dv = -\int_{r=1} \mathbf{x}\mathbf{x}_r\, d\theta + r' \int_{r=r'} \mathbf{x}\mathbf{x}_r\, d\theta$$

where $r'$ is the radius of $\rho'$. Hence

$$-\int_{r=1} \mathbf{x}\mathbf{x}_r\, d\theta < A_1 + r' \int_{r=r'} |\mathbf{x}|\ |\mathbf{x}_r|\, d\theta.$$

By (1)

$$\int_{r=1} |\mathbf{x}_r|\, d\theta < AA_1 + Ar' \int_{r=r'} |\mathbf{x}|\ |\mathbf{x}_r|\, d\theta$$

or because of the boundedness of $|\mathbf{x}|$ and $|\mathbf{x}_r|$ on $\bar{\rho}$,

$$\int_{r=1} |\mathbf{x}_r| \, d\theta < A_2 \tag{2}$$

where $A_2$ is a constant. Now, $\mathbf{x}$ representing a minimal surface, we have on $r = 1$ the relation $|\mathbf{x}_r| = |\mathbf{x}_\theta|$. Hence

$$L = \int_{r=1} |\mathbf{x}_\theta| \, d\theta < A_2 \tag{3}$$

where $L$ is the length of $\rho$ and $A_2$ is a constant. The statement (3) and consequently the conclusion (4) is thus proved.

It remains to establish beyond ($a$) the assertion (2) for our polyhedra $K$: the circle $\bar{\rho}$ is divided into a finite number of arcs such that each arc is mapped by $\mathbf{x}(u, v)$ either on a segment of an edge, or onto a curve on a face (that the mapping is analytic on these arcs follows from the principle of reflection).

To complete the proof of statement (2) a further chain of arguments is needed.[8]

First some preliminary remarks: In an arbitrary neighborhood of a bridge $\gamma$ of diameter less than $\epsilon$ there are bridges of diameter less than $\epsilon$ whose end-points are analytic; for, points on $\bar{\rho}$ where the (continuous) vector $\mathbf{x}$ is not analytic are nowhere dense and the basic lemma can be applied to bridges $\gamma$ with end-points in intervals of analyticity.

Second: Lines $x' = ax + by + cz + d = 0$ with constant $a, b, c, d$ are equipotential lines in $\bar{M}$; curves $x' = 0$ in $\bar{M}$ which separate domains $x' > 0$ from $x' < 0$ are unambiguously defined.

Third: A bridge $\bar{\gamma}$ analytic including its end-points can have only a finite number of intersections with curves $x' = 0$ unless $\bar{\gamma}$ itself is a curve $x' = 0$ and hence $\gamma$ a plane curve.

Now we introduce the concept of a plane $T$ *transversal*[9] to $K$ at a point $p$: A plane $T$ through $p$ on $K$ with trace $\tau$ on $K$ is called transversal if for each bridge[10] $\beta$ sufficiently near to $p$ and lying in $T$ the plane part of $T$ between $\beta$ and $\tau$ is the surface of least area bounded by $\beta$ and $K$. Obviously for $p$ on a face $A$ each plane $T$ through $p$ perpendicular to $A$ is a transversal plane. Furthermore, for $p$ on an edge $\alpha$ each plane through $\alpha$ normal to a supporting plane through $\alpha$ is transversal; finally for a vertex $p$ there exists at least one edge $\alpha$ such that the plane $T$ through $\alpha$ and perpendicular to the opposite face $A$ of the trihedral angle in $p$ is transversal. (This latter statement is a non-trivial consequence of the assumption that all the angles at the edges of $K$ are obtuse.)

To refute the assumption that statement (2) is wrong the transversal planes now are used in the following manner: Suppose first that the trace $\rho$ (continuous by theorem ($a$)) intersects an edge $\alpha$ infinitely often in the neighborhood of a point $p$ of $\alpha$, $p$ not a vertex. Let $x' = 0$ be a transversal plane $T$ through $p$, let $\bar{\gamma}$ and $\gamma$ be a small bridge (original and image) near

$\bar{p}$ or $p$; consider a plane curve $x' = 0$ separating a zone $x' > 0$ from a zone $x' < 0$ on $M(\gamma)$ and the corresponding equipotential line in $\bar{M}$. The latter, starting from $\bar{p}$ must end on $\bar{\gamma}$; for otherwise it would form a small bridge $\bar{\beta}$ and consequently $M(\beta)$ would be a part of the plane $T$, and $M$ would be altogether plane, which can be excluded from the outset. Since infinitely many successive intervals $x' > 0$ and $x' < 0$ on $\bar{p}$ near $\bar{p}$ are assumed we would have infinitely many points on $\bar{\gamma}$ with $x' = 0$, hence $\bar{\gamma}$ would be a line $x' = 0$ which can be excluded by proper choice of $\bar{\gamma}$ unless $M$ is plane.

In a similar way we can rule out the possibility that a point $p$ on $\alpha$ is a point of accumulation of points where $\rho$ reaches $\alpha$ from a face $B$ without crossing. In this case we use transversal planes $T(d)$ normal to $B$ at the distance $d$ from $\alpha$ and then let $d$ tend to zero whereupon a similar contradiction as above is obtained.

If $\bar{p}$ on $\bar{p}$ is mapped on a vertex $p$ and if infinitely many points on $\bar{p}$ near $\bar{p}$ are mapped on $p$ or on points on edges through $p$ while infinitely many other intermediate points are mapped on points in faces, a contradiction is obtained by using the transversal plane $x' = 0$ through $p$ in a way not essentially different from that sketched above.

Combined, these results establish the assertion (2) and thus complete the proof of our theorem.

[1] See, e.g., Courant, "The Existence of Minimal Surfaces under Prescribed Boundary Conditions," *Acta Math.*, **72**, 51–97 (1940), in particular, p. 81 ff.

[2] See loc. cit., p. 95.

[3] The brief presentation here will be amplified in a book on *Dirichlet's Principle, Conformal Mapping and Plateau's Problem.*

[4] Naturally, in all the various cases of Plateau's problem with free boundaries similar theorems and proofs apply; moreover, the methods used here indicate the possibility of attacking more general variational problems with free boundaries in more than one dimension.

The author wishes to acknowledge that discussions and correspondence with Dr. C. L. Siegel of the Institute for Advanced Study have played a decisive part in carrying out the present investigation.

[5] See loc. cit., p. 70.

[6] Quite generally any piece of a surface, if projected along the outward normal onto a non-negatively curved surface, will be reduced in area. This remark could be used to generalize the theorem (*b*) by substituting for the cone other non-negatively curved surfaces.

[7] The proof follows from a simple geometrical construction: If $M$ ends in a piece $\alpha$ of an edge and forms an acute angle with an adjacent face $A$, one could substitute for $M$ an admissible surface of smaller area by replacing a part of $M$ adjacent to $\alpha$ and otherwise bounded by an analytic arc $\lambda$ joining the end points of $\alpha$ and drawn sufficiently near to $\alpha$ by the cylindrical surface projecting $\lambda$ onto $A$.

[8] Details will be given in the forthcoming book.

[9] The concept and its application seem to be capable of generalizations to boundary surfaces $S$ not plane, and to other variational problems.

[10] The term "bridge" is used here in a slightly generalized way meaning any Jordan arc $\beta$ outside $K$ joining two points on $K$.

# THE LAPLACE EQUATION IN SPACE

BY EDWARD KASNER AND JOHN DE CICCO

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY AND ILLINOIS INSTITUTE OF TECHNOLOGY

Communicated June 29, 1945

1. A function $h(x, y)$ is harmonic if it satisfies the Laplace equation in two dimensions $h_{xx} + h_{yy} = 0$. The one-parameter family of curves $h(x, y) = c$ is called an isothermal family and $c$ is said to be the isothermal parameter. A family of curves $f(x, y) = c$ is isothermal if and only if $f$ is a function of a harmonic function, and, therefore, if $f$ satisfies a certain partial differential equation of third order.

Sophus Lie obtained the first intrinsic characterization of the set of isothermal families, which may be stated in the following form. A family of $\infty^1$ curves is isothermal if and only if the inclination $\theta$ is a harmonic function of $(x, y)$. From this it follows that the angle between any two isothermal families is a harmonic function of $(x, y)$.

The authors have investigated all possible types of transformations preserving the class of isothermal families. We have found extensive classes of differential element transformations converting isothermal families into isothermal families. For the case of lineal-element transformations, the following result has been proved. The group of lineal-element transformations preserving the class of isothermal families is: $X = \phi(x, y)$, $Y = \psi(x, y)$, $\Theta = a\theta + h(x, y)$, where $\phi$ and $\psi$ satisfy the direct or reverse Cauchy-Riemann equations: $\phi_x = \pm\psi_y$, $\phi_y = \mp\psi_x$, $a$ is a non-zero constant and $h$ is any harmonic function of $(x, y)$. By this, we establish the theorem that the only point transformations which send every isothermal family into an isothermal family, are the conformal. (This is more general and more difficult than the obvious problem of finding the auto-transformations of the Laplace equation.)

2. In our present work, we shall study the case in three dimensions, where the situation is entirely different both analytically and geometrically. A function $h(x, y, z)$ is harmonic if it satisfies the Laplace equation $h_{xx} + h_{yy} + h_{zz} = 0$. The $\infty^1$ surfaces $h(x, y, z) = c$ is called an isothermal family and $c$ is said to be the isothermal parameter. A family of surfaces $f(x, y, z) = c$ is isothermal if and only if $f$ is a function of a harmonic function, and, therefore, if and only if $f$ satisfies two partial differential equations of third order. Now we state our new theorems.

THEOREM 1. *If $h$ is any harmonic function of three variables $(x, y, z)$ with $h_x \neq 0$, $h_y \neq 0$, $h_z \neq 0$, then a function $f = f(h_x, h_y, h_z)$ is harmonic for any arbitrary $h$ only when $f$ is linear integral in the three arguments.*

This theorem can be extended to partial derivatives of $h$ of any order,

that is, the function $f$ must be linear integral in all the partial derivatives. There is a similar theorem which is valid for any dimension $n \geqq 3$.

However, for $n = 2$ the situation is quite different. If $h$ is a harmonic function of two variables $(x, y)$, with $h_x \neq 0$, $h_y \neq 0$, then $f = f(h_x, h_y)$ is harmonic for any $h$ if and only if $f$ is harmonic in the two arguments. Thus arc tan $h_y/h_x$, log $(h_x^2 + h_y^2)$, $e^{h_x} \cos h_y$, $h_x^2 - h_y^2$ are all harmonic functions for any harmonic function $h$, although non-linear.

3. THEOREM 2. *Lie's characterization of isothermal families in the plane is not valid in space. That is, the angle between any isothermal system of surfaces and a set of parallel planes is not necessarily a harmonic function.*

This result can be deduced from Theorem 1. For if $h(x, y, z)$ is a harmonic function, then the angle $\gamma$ between the isothermal system of surfaces $h(x, y, z) =$ const. and the pencil of parallel planes $z =$ const., is a complicated expression involving the partial derivatives of $h$. By Theorem 1, this angle $\gamma$ cannot be a harmonic function for every $h$.

4. THEOREM 3. *The system of simultaneous partial differential equations of first order:* $\partial z/\partial x = p(x, y, z)$, $\partial z/\partial y = q(x, y, z)$, *possesses as solution an isothermal family of surfaces if and only if $p$ and $q$ satisfy the set of three partial differential equations of second order.*

$$
\begin{aligned}
p_y + qp_z &= q_x + pq_z, \\
\tfrac{1}{2}(1 + p^2 + q^2)(p_{xx} + p_{yy} + p_{zz}) &= (p_y q_z - p_z q_y) + p(p_x^2 + p_y^2 + p_z^2 + \\
&\quad p_x q_y - p_y q_x) + q(p_x q_z + p_y q_y + p_z q_z), \\
\tfrac{1}{2}(1 + p^2 + q^2)(q_{xx} + q_{yy} + q_{zz}) &= (p_x q_z - p_z q_x) + q(q_x^2 + q_y^2 + q_z^2 + \\
&\quad p_x q_y - p_y q_x) + p(p_x q_z + p_y q_y + p_z q_z). \qquad \text{(I)}
\end{aligned}
$$

The first of these conditions is the condition of integrability.

As an application of Theorem 3, the following result may be established. The only families of $\infty^1$ planes which are isothermal are the pencils of planes. For straight lines the analogous result was proved by Lagrange. The proof in space is difficult.

5. THEOREM 4. *The only point transformations converting every isothermal system of surfaces into an isothermal system are those of the Liouville inversive group.*

This proposition is established with the aid of Liouville's theorem which states that the conformal group of space is exactly the inversive group of ten parameters.

6. At this stage, let us consider transformations of surface elements $(x, y, z, p, q)$. The Lie contact group satisfies the conditions

$$
\left.
\begin{aligned}
Z_x + pZ_y &= P(X_x + pX_z) + Q(Y_x + pY_z), \; Z_p = PX_p + QY_p, \\
Z_y + qZ_z &= P(X_y + qX_z) + Q(Y_y + qY_z), \; Z_q = PX_q + QY_q.
\end{aligned}
\right\} \qquad \text{(L)}
$$

The Lie group has the property that it converts every union, that is, any

double series: $z = z(x, y)$, $p = p(x, y)$, $q = q(x, y)$, with the property that $dz = pdx + qdy$, into a union.

Kasner studied transformations of surface elements which carry every integrable field of planar elements, that is, $p = p(x, y, z)$, $q = q(x, y, z)$, which satisfy the condition $p_y + qp_z = q_x + pq_z$, into an integrable field. It is not assumed that a union of an integrable field is carried into the corresponding union of the associated integrable field. Kasner proved that *this group of transformations is identical with the Lie contact group.*

7. By means of Kasner's theorem, it follows that any transformation of surface elements which carries every isothermal field into an isothermal field must be a contact transformation. Then the following result may be established.

THEOREM 5. *The only transformations of surface elements which send every isothermal field into an isothermal field are those of the Liouville inversive group.*

Thus in space the only possible transformations preserving the isothermal character must be point transformations generated by inversions with respect to a sphere.

[1] Kasner, *Differential Geometric Aspects of Dynamics*, Princeton Colloquium Lectures, 1913, 1934.

[2] Kasner, "A Characteristic Property of Isothermal Families," *Mathematische Annalen*, **59**, 352–354 (1904).

[3] De Cicco, "New Proofs of the Theorems of Beltrami and Kasner on Linear Families of Curves," *Bull. Am. Math. Soc.*, **49**, 407–412 (1943).

[4] De Cicco, "The Two Conformal Covariants of a Field," *Revista de Matemáticas de Tucumán*, **2**, 59–66 (1941).

[5] Kasner, "Lineal-element Transformations Which Preserve the Isothermal Character," these PROCEEDINGS, **27**, 406–412 (1941).

[6] Kasner, "Transformation Theory of Isothermal Families and Certain Related Trajectories," *Revista de Matemáticas y Física Teórica de la Universidad de Tucumán*, **2**, 17–24 (1941), Argentina.

[7] Kasner and De Cicco, "Generalized Transformation Theory of Isothermal and Dual Families," these PROCEEDINGS, **28**, 52–55 (1942).

[8] Kasner and De Cicco, "An Extensive Class of Transformations of Isothermal Families," *Revista de Matemáticas y Física Teórica de la Universidad de Tucumán*, **3**, 271–282 (1942).

[9] Kasner and De Cicco, "Transformation Theory of Isogonal Trajectories of Isothermal Families," these PROCEEDINGS, **28**, 328–333 (1942).

[10] Kasner, "Geometric Properties of Isothermal Families," *Facultad de Ciencias Matemáticas de la Universidad Nacional del Litoral*, **5**, 1–10 (1943), Rosario, Argentina.

[11] Kasner, "Lineal-Element Transformations of Space for Which Normal Congruences of Curves are Converted into Normal Congruences," *Duke Mathematical Journal*, **5**, 72–83 (1939).

# ALGEBRAIC CURVES, SYMMETRIES AND SATELLITES

BY EDWARD KASNER

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Communicated June 29, 1945

Schwarz defined symmetry with respect to any real analytic base curve. This is known as Schwarzian reflection and is used extensively in the theory of analytic prolongation of functions of a complex variable. The author defined the process geometrically and intrinsically in the Proceedings of the Cambridge International Congress (1912) and called it *conformal symmetry*. If the base curve is denoted by $C$, then the image of any point $P$ of the plane is obtained in the following manner. Through $P$, draw the two minimal lines, which will intersect $C$, respectively, in the two points $Q_1$ and $Q_2$. The remaining minimal lines passing through $Q_1$ and $Q_2$ will intersect in a point $P'$. This point $P'$ is the image of the point $P$ with respect to the base curve $C$. *Conformal symmetry is the only reverse conformality which leaves the individual points of C fixed. No other reverse conformalities of period two can exist.*

In the present paper, I wish to discuss the case where the base curve $C$ is a general algebraic curve. We thus operate in the complete complex (four-dimensional) plane. Schwarz was considering only the real (two-dimensional) gaussian plane.

In the Schwarzian theory, the transformation is one valued because only the local neighborhood of the base curve $C$ is considered. But in the present paper, since we are considering the total plane, the transformation $T$ is many valued. If the algebraic curve $C$ is of degree $n$, then the transformation $T$ is always algebraic and is, in general, of degree $n^2$.

If, in particular, the base curve $C$ is a conic, the transformation $T$ will convert, in general, one point $P$ into four points $P'$. But if the conic is a circle, we obtain ordinary inversion, which is of course one valued. There are mixed imaginary cases of conics where the degree is neither four nor one but is actually two.

Usually we say that the image of the base curve $C$ with respect to itself is $C$. But in the algebraic case we are now considering, this is no longer exactly true. The complete image of $C$ consists partly of $C$ and partly of a new curve which we define as the *satellite* of $C$ and denote by $S$. The image of a random curve $C'$ of degree $n$ with respect to $C$ is, in general, of degree $n^2$. But the image of $C$ with respect to itself is reducible, and we find that one branch is the curve $C$ counted a certain number of times, and the new satellite curve $S$.

In the case of a conic $C$, the satellite $S$ is another conic. The new conic $S$ is confocal with the old conic $C$, and has uniquely determined diameters.

If $C$ is a rectangular hyperbola, the satellite conic $S$ is identical with $C$. Of course, the circle has no satellite.

If we consider the totality of $\infty^5$ conics of the plane, the induced transformation from each conic to its satellite is of the fifth degree (in the coefficients of the conic). This induced transformation is neither a point transformation nor a contact transformation. Of course, all the contact transformations carrying the set of $\infty^5$ conics into itself form merely the total projective group consisting of collineations and correlations.

THEOREM 1. *The degree of the satellite $S$ of a general algebraic curve $C$ of degree $n$ is $n(n-1)^2$.*

This is true only in general since actually the degree of the satellite may in special cases be lower. From this formula, we see that the satellite of a conic is a conic; but the satellite $S$ of a cubic curve $C$ is, in general, an algebraic curve of degree 12.

A noteworthy case where the degree of the satellite is lower is stated in the following result.

THEOREM 2. *The satellite $S$ of every algebraic potential curve $C$ is the curve $C$ itself.*

The curve $\phi(x, y) = 0$, where $\phi$ is a polynomial of degree $n$ in $(x, y)$, is called a potential curve if $\phi$ is harmonic, that is, $\phi_{xx} + \phi_{yy} = 0$. In this case, although the transformation $T$ associated with the curve is of degree $n^2$, reducibility studies show that the satellite is of exceptionally low degree. When the degree is 2, we obtain the previously noted example of the equilateral hyperbola.

We have already observed that if the degree of $C$ is $n$, the degree of $T$ is $n^2$. What then is the degree of $T^2$? We might expect it to be of degree $n^4$. But we prove the following proposition (on account of reducibility).

THEOREM 3. *The degree of $T^2$ is $n^2(n-1)^2$.*

All this theory depends on the fact that the transformations are algebraic multivalued, and that reducibility phenomena occur.

THEOREM 4. *If the base curve $C$ is a conic, $T$ is of degree 4 as already noted; and we now state that $T^2$ is also of degree 4, and all the iterations of $T$ give transformations of degree 4.*

It thus turns out that the various powers of $T$ form a discontinuous set. This set has the combinatorial property but no power of $T$ is the identity. Thus this set, although closed, cannot be a group.

An interesting case occurs in the imaginary domain when the base conic $C$ goes through only one circular point at infinity. The transformation $T$ is then of degree 2. We find that all the powers of $T$ are of degree 2. It results that $T^3 = T$ but no power of $T$ is the identity. Thus, in this case, the complete set of all powers of $T$ consists essentially of only the two transformations $T$ and $T^2$.

We also have obtained results on *related families* of curves $\phi(x, y) = C$

their satellites, and set of symmetries, and a remarkable *induced transformation* of lower degree than the symmetries. For the case of conics the induced transformation is linear, and is in fact an affinity. This affinity enables us to construct the satellite conic easily in the real domain.

[1] Kasner, "Algebraic Potential Curves," *Bull. Am. Math. Soc.*, 1901.

[2] Kasner, "Conformal Geometry," *Proc. Fifth Internat. Cong. of Math.*, **2**, 81 (1912).

[3] Kasner, "Geometry of Conformal Symmetry (Schwarzian Reflection)," *Ann. Math.*, **38**, 873 (1937).

[4] Comenetz, "Conformal Geometry on a Surface," *Ibid.*, **39**, 863 (1938).

I wish to thank J. De Cicco, Aida Kailish, F. Supnik and C. Vlucay for assistance in elaborating the theory of satellites.

---

## *DIFFERENTIAL EQUATIONS IN FRÉCHET DIFFERENTIALS OCCURRING IN INTEGRAL EQUATIONS*

BY ARISTOTLE D. MICHAL

CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated July 5, 1945

*Introduction.*—This paper is concerned with the study of the resolvent kernels and solutions of Volterra and Fredholm integral equations as functionals of the given kernels. A study is made of the completely integrable equations in Fréchet differentials characterizing these functionals of the kernels. The investigations have immediate application to the problem of obtaining approximations to the resolvent kernels and the solutions of integral equations with precise estimates of the errors. The above results are applied to differential corrections for some differential equations. *As an important by-product we obtain a solution of a century-old problem in non-commutative analysis.* We assume that the reader is conversant with the fundamental theorems on Fréchet differentials of functionals.

*1. Resolvent Kernels and Solutions of Volterra Integral Equations as Functionals of the Kernels.*—Consider the Volterra integral equation of the second kind

$$f(x) = y(x) + \int_a^x K(x, \xi)y(\xi)d\xi (T: a \leq \xi \leq x \leq b) \qquad (1)$$

with continuous kernel in $T$ and continuous $f(x)$ in $(a, b)$. The well-known unique continuous solution $y(x)$ of equation (1) in the closed interval $(a, b)$ is

$$y(x) = f(x) + \int_a^x k(x, \xi)f(\xi)d\xi, \qquad (2)$$

where the resolvent kernel $k(x, \xi)$ has the expansion

$$k(x, \xi) = -K(x, \xi) + K^2 - K^3 + \ldots + \ldots \qquad (3)$$

in terms of the composition powers of the kernel $K(x, \xi)$:

$$K^2 = \int_\xi^x K(x, s)K(s, \xi)ds, \qquad K^{i+1} = \int_\xi^x K(x, s)K^i(s, \xi)ds. \tag{4}$$

We shall write $k(x, \xi)$ as $k[K]$ to show its functional dependence on the kernel $K(x, \xi)$. Since $k[K]$ is an entire analytic functional of $K(x, \xi)$, it follows from a theorem[1] of the author on the Fréchet differentiability of regular power series in normed linear spaces that the Fréchet differential $\delta k[K]$ of $k[K]$ with increment $\delta K$ exists for each continuous $K(x, \xi)$ in $T$ and is given by a term by term Fréchet differentiation of (3). The result written in terms of composition products is

$$\delta k[K] = -(\delta K + k\delta K + \delta K k + k\delta K k). \tag{5}$$

For example $k\delta K$ stands for $\int_\xi^x k(x, s)\delta K(s, \xi)ds$. In calculating this Fréchet differential and in considering $k[K]$ as an analytical functional of $K(x, \xi)$, we understand that the independent and dependent variables in $k[K]$ are in the complete normed linear space (a Banach space) of all continuous $K(x, \xi)$ in $T$ with norm defined by

$$||K|| = \max_{x,\, \xi \in T} |K(x, \xi)|. \tag{6}$$

In (5), $K$ as well as the increment $\delta K$ are arbitrary continuous functions of $x$ and $\xi$ in $T$.

Another way of obtaining formula (5) is to use the existence of the Fréchet differential $\delta k[K]$, guaranteed by the author's theorem on regular power series in normed linear spaces, and then to *compute* it from either of the classical resolvent relations of integral equation theory:

$$K + k + Kk = 0, \qquad k + K + kK = 0. \tag{7}$$

In fact, if we take the Fréchet differential of both sides of the first of these identities, we obtain

$$\delta K + \delta k + \delta K k + K\delta k = 0. \tag{8}$$

It can be shown that the unique continuous solution of this integral equation (8) with $\delta k$ as unknown is given precisely by (5).

Since $\delta k$ exists given by (5), the Fréchet differential of the solution $y[K/x]$ of the integral equation (1) considered as a functional of the kernel $K(x, \xi)$ can be computed from (2). A simpler derivation is the following which does not use the result (5) but merely the existence of $\delta k$ and hence of $\delta y[K/x]$ from (2). Take the Fréchet differential of both sides of (1) and obtain

$$0 = \delta y[K/x] + \int_a^x \delta K(x, \xi)y[K/\xi]d\xi + \int_a^x K(x, \xi)\delta y[K/\xi]d\xi.$$

If we solve this Volterra integral equation of the second kind for $\delta y[K/\xi]$, we obtain a result which can be reduced to the form

$$\delta y[K/x] = -\int_a^x \{\delta K(x, \xi) + \int_\xi^x k(x, s)\delta K(s, \xi)ds\}y[K/\xi]d\xi. \tag{9}$$

In the above discussion it is understood that the Fréchet differential $\delta y[K/x]$ is defined for a functional $y[K/x]$ with values in the well-known Banach space of continuous functions of $x$. Results (5) and (9) for one Volterra integral equation can, by an evident change of interpretation of notation, be shown to hold for a system of $n$ Volterra integral equations of the second kind in $n$ unknowns. If the system of equations is written as (1) with $f(x)$ and $y(x)$ as column matrices of functions of $x$ and $K(x, s)$ as the $n$-rowed square matrix of the $n^2$ kernels, then in (2), $k(x, \xi)$ will stand for the $n$-rowed square matrix of the $n^2$ resolvent kernels. In (5) products are to be interpreted as combined matric and integral composition products of the first kind; for example, $k\delta K$ will stand for $\int_\xi^x k_\alpha^i(x, s)\delta K_j^\alpha(s, \xi)ds$ with the summation convention operating on indices. In (9), products of functions will be matric products.

*2. The Completely Integrable Total Differential Equations in Fréchet Differentials of Functionals.*—Let us now return to the one Volterra integral equation (1) and inquire into the properties of the functional equations (5) and (9) satisfied, respectively, by the resolvent kernel and the solution of (1) as functionals of the kernel $K(x, \xi)$.

It can be shown from (5) with the aid of the known theorems on differentials and a mathematical induction that all successive Fréchet differentials of the resolvent kernel $k[K]$ exist at any chosen continuous kernel $K(x, \xi)$ in $T$. If $\delta_1 K(x, \xi)$, $\delta_2 K(x, \xi)$, ..., $\delta_n K(x, \xi)$ are the increment functions for the $n$th successive Fréchet differential $\delta_n\delta_{n-1}\ldots\delta_1 k[K]$, we find that

$$\left.\begin{aligned}\delta_n\delta_{n-1}\ldots\delta_1 k[K] = (-1)^n P_{1,2,\ldots,n}\{[\delta_1 K + k\delta_1 K)(\delta_2 K + \\ k\delta_2 K)\ldots(\delta_n K + k\delta_n K) + (\delta_1 K + k\delta_1 K)(\delta_2 K + k\delta_2 K)\ldots \\ (\delta_n K + k\delta_n K)k\}(n = 1, 2, 3, \ldots),\end{aligned}\right\} \quad (10)$$

where $P_{1,2\ldots,n}$ stands for the sum of $n!$ terms obtained from (and including) the bracket by a permutation of the integers $1, 2, \ldots, n$. If all increments are equal, the $n$th Fréchet differential $\delta^n k[K]$ is given by

$$\delta^n k[K] = (-1)^n n!\{(\delta K + k\delta K)^n + (\delta K + k\delta K)^n k\}(n = 1, 2, 3, \ldots). \quad (11)$$

The total differential equations (5) and (9) are *completely integrable* since the condition for integrability is identically satisfied in all continuous $K$, $k$, $\delta_1 K$, $\delta_2 K$ (considered as independent variables) in $T$. Although equations (5) and (9) are completely integrable for all continuous kernels $K$ in $T$, the existence and uniqueness theorems[2] in normed linear spaces of Michal and Elconin are not strong enough to insure the unicity of the solution taking on arbitrary initial conditions. We can, however, give an argument which will furnish uniqueness theorems for functional equations (5) and (9). In fact let $\lambda[K]$ satisfy the following equation in Fréchet differentials $\delta\lambda[K]$ for all kernels $K(x, \xi)$ continuous in $T$:

$$\delta\lambda[K] = -(\delta K + \lambda\delta K + \delta K\lambda + \lambda\delta K\lambda), \tag{12}$$

and let it take on the initial condition $\lambda[0] = 0$, the identically vanishing function in $T$. We shall find the most general such entire analytical functional.[3] Assume then that

$$\lambda[K] = \sum_{i=1}^{\infty} \lambda_i[K] \tag{13}$$

so that $\lambda_i[K]$ is a homogeneous functional polynomial of degree $i$. It follows from R. S. Martin's contributions to abstract polynomials[4] that to each homogeneous polynomial $\lambda_i[K]$ there exists a unique polar, i.e., a completely symmetric multilinear functional $\Lambda_i[K_1, K_2, \ldots, K_i]$ such that $\Lambda_i[K, K, \ldots, K] = \lambda_i[K]$. It is readily shown that the Fréchet differential $\delta\lambda_i[K]$ is given by

$$\delta\lambda_i[K] = i\Lambda_i[K, K, \ldots, K, \delta K]. \tag{14}$$

By the author's theorem[1] on the Fréchet differentiability of power series in normed linear spaces, we have

$$\delta\lambda[K] = \sum_{i=1}^{\infty} i\Lambda_i[K, K, \ldots, \delta K] \tag{15}$$

for all $K(x, \xi)$ continuous in $T$.

On using (13), (14) and (15) in (12), one can show that $\lambda_1[K] = -K$, $\lambda_2[K] = K^2$ and that the following recurrence relations hold for $n \geq 2$:

$$(n+1)\Lambda_{n+1}[K, K, \ldots, K, \delta K] = -\lambda_n[K]\delta K - \delta K\lambda_n[K] - \sum_{\substack{i+j=n \\ i,j\geq 1}} \lambda_i[K]\delta K\lambda_j[K].$$

From these results one finds that

$$\lambda[K] = -K + K^2 - K^3 + \ldots,$$

and hence from our earlier results on resolvent kernels the fundamental

THEOREM 1. *The completely integrable non-linear total differentia system in Fréchet differentials*

$$\left.\begin{aligned} \delta\lambda[K] &= -(\delta K + \lambda\delta K + \delta K\lambda + \lambda\delta K\lambda) \\ \lambda[0] &= 0 \end{aligned}\right\} \tag{16}$$

*has a unique entire analytic functional solution given by*

$$\lambda[K] = -K + K^2 - K^3 + \ldots,$$

*the resolvent kernel of the arbitrary continuous kernel $K(x, \xi)$ in $T$.*

By similar methods one can prove the following theorem.

THEOREM 2. *The completely integrable linear total differential system in Fréchet differentials, with $k[K/x, \xi]$ the resolvent kernel of $K(x, \xi)$,*

$$\left.\begin{aligned} \delta Z[K/x] &= -\int_a^x \{\delta K(x, \xi) + \int_\xi^x k[K/x, s]\delta K(s, \xi)ds\} Z[K/\xi]d\xi \\ Z[0/x] &= f(x) \end{aligned}\right\} \quad (17)$$

*has a unique entire analytic functional solution given by*

$$Z[K/x] = f(x) + \int_a^x k[K/x, \xi]f(\xi)d\xi, \quad (18)$$

*the solution of the Volterra integral equation* (1).

In this theorem we understand that the definitions of Fréchet differentials and analytic functionals for $Z[K/x]$ are given with the independent variable $K(x, \xi)$ ranging over the Banach space of the previous theorem while the value space is the Banach space of functions of $x$ continuous in the interval $a \leq x \leq b$ and having as norm the maximum of the absolute value of the function over the interval. The details of proof are facilitated by writing the differential system as

$$\begin{aligned} \delta Z[K] &= -(\delta K + k[K]\delta K)\cdot Z[K] \\ Z[0] &= f \end{aligned}$$

and by writing $A \cdot w$ for the bilinear functional $\int_a^x A(x, \xi)w(\xi)d\xi$.

A more general theorem can be proved by different methods on making use of Theorem 1 and the result $\delta(Z[K] + K\cdot Z[K]) = 0$.

THEOREM 3. *There is one and only one solution of the completely integrable system (17) in Fréchet differentials. It is given by (18).*

*3. Simple Illustrative Applications to Ordinary Differential Equations.*— The differential system ($P(x)$ and $Q(x)$ continuous, say, in $(a, b)$)

$$\frac{dy(x)}{dx} + P(x)y(x) = Q(x), \qquad y(a) = y_0 \quad (19)$$

is equivalent to an integral equation (1) with kernel $K(x, \xi) = P(\xi)$ and known function $f(x) = y_0 + \int_a^x Q(s)ds$. With the aid of Theorem 2 one can show that the solution $y[P/x]$ of the system (19) as a functional of $P(x)$ is the unique entire analytic functional solution of the total differential system in Fréchet differentials.

$$\left.\begin{aligned} \delta y[P/x] &= -\int_a^x e^{-\int_s^x P(u)du} y[P/s]\delta P(s)ds \\ y[0/x] &= y_0 + \int_a^x Q(t)dt. \end{aligned}\right\} \quad (20)$$

As another simple example we take the differential system

$$\frac{d^2y(x)}{dx^2} + P(x)y(x) = 0, \qquad y(a) = y_0, \quad \left.\frac{dy(x)}{dx}\right]_{x=a} = y_1$$

($P(x)$ continuous in $(a, b)$). (21)

This is equivalent to an integral equation (1) with kernel $K(x, \xi) = (x - \xi) P(\xi)$ and known function $f(x) = y_0 + y_1(x - a)$. Again with the aid of Theorem 2 one can show that the solution $y[P/x]$ of the system (21) as a functional of $P(x)$ is the unique entire analytic functional solution of the differential system in Fréchet differentials

$$\left.\begin{aligned} \delta y[P/x] &= \int_a^x \mu[P/x, \xi] y[P/\xi] \delta P(\xi) d\xi \\ y[0/x] &= y_0 + y_1(x - a), \end{aligned}\right\} \tag{22}$$

where

$$\mu[P/x, \xi] = -(x - \xi) - \int_\xi^x F[P/x, \eta] (\eta - \xi) d\eta \tag{23}$$

and $F[P/x, \eta]$ is the resolvent kernel of the kernel $K(x, \eta) = (x - \eta)P(\eta)$.

4. *The Matric Exponential in Non-Commutative Analysis.*—The results of Theorem 1, Theorem 2 and Theorem 3 can easily be extended to the case of a system of $n$ Volterra integral equations. We shall apply such a generalized Theorem 2 to the system of differential equations with constant coefficients

$$\frac{dZ^i(x)}{dx} = a^i_j Z^j \qquad Z^i(0) = Z^i_0. \tag{24}$$

This is equivalent to the matric Volterra integral equation

$$Z_0 = Z(x) + \int_0^x K(x, \xi) Z(\xi) d\xi \tag{25}$$

where $K(x, \xi) = -A = -||a^i_j||$, the matrix of the constants $a^i_j$.

If we write the solution of (25) as $Z[K/x]$, we see that

$$Z[-A/1] = e^A Z_0, \tag{26}$$

where $e^A$ is the *matric exponential.* With the aid of the generalized Theorem 2, we can show that

$$\begin{aligned} &\delta(e^A Z_0) = \int_0^1 \left\{ \delta A + \int_\xi^1 A e^{(1-s)A} \delta A ds \right\} e^{\xi A} Z_0 d\xi, \\ &k[0] = 0. \end{aligned}$$

This leads to the following important result giving the functional equation satisfied by the matric exponential.

THEOREM 4. *The matric exponential function $Z(A) = e^A$, satisfies the folowing system in Fréchet differentials:*[5]

$$\left.\begin{aligned} \delta Z(A) &= \int_0^1 Z((1 - \xi)A) \delta A Z(\xi A) d\xi \\ Z(0) &= I, \textit{ the unit matrix.} \end{aligned}\right\} \tag{27}$$

5. *Resolvent Kernels and Solutions of Fredholm Integral Equations as Functionals of the Kernels.*—The set $\Delta$ of kernels $K(x, \xi)$ continuous in the square $S: a \leq x, \xi \leq b$ and with non-vanishing Fredholm determinants forms an open set in the Banach space of all continuous $K(x, \xi)$ in $S$. If we now write a composition of the second kind (constant limits) merely as a product, the following result can be proved.

THEOREM 5. *The Fréchet differential of the resolvent kernel $k[K]$ of a kernel $K(x, \xi)$ exists at each $K(x, \xi) \epsilon \Delta$, and for all $K(x, \xi) \epsilon \Delta$ it satisfies the differential system in Fréchet differentials*

$$\delta k[K] = -(\delta K + k\delta K + \delta K k + k\delta K k),$$
$$k\,[0] = 0.$$

The proof of this theorem[6] differs from the power series proof for the corresponding result in Volterra integral equations.

The correspondent of Theorem 3 can be proved by entirely analogous methods. We shall state it as

THEOREM 6. *There is one and only one solution of the completely integrable system in Fréchet differentials*

$$\delta Z[K/x] = -\int_a^b \{\delta K(x, \xi) + \int_a^b k[K/x, s]\delta K(s, \xi)ds\} Z[K/\xi]d\xi,$$
$$Z[0/x] = f(x), \qquad (K \epsilon \Delta)$$

*where $k[K/x, s]$ is the resolvent kernel of $K(x, \xi)$. It is given by the solution of the Fredholm integral equation with kernel $K(x, \xi)$ and known function $f(x)$.*

[1] Michal, A. D., *The Fréchet Differentials of Regular Power Series in Normed Linear Spaces*. To be published elsewhere.

[2] Michal, A. D., and Elconin, V., "Completely Integrable Differential Equations in Abstract Spaces," *Acta Mathematica*, **68**, 71–107 (1937).

[3] In Banach spaces, $\sum_{i=1}^{\infty} \lambda_i[K]$ defines an entire analytic function if the corresponding real power series $\sum_{i=1}^{\infty} m_i x^i$ converges for all real $x$. By $m_i$ we mean the modulus of the homogeneous polynomial $\lambda_i[K]$. For analytic functions in Banach spaces see Michal, A. D., and Martin, R. S., "Some Expansions in Vector Space," *Jour. Math. Pures et Appl.*, **13**, 69–91 (1934). Presented to the American Mathematical Society, Sept., 1932. See also Martin, R. S., *Contributions to the Theory of Functionals*, a California Institute of Technology Ph.D. thesis (unpublished), 1932, written under the direction of the present author. Some portions of this thesis have trickled into the later mathematical literature.

[4] Martin, R. S., *Contributions to the Theory of Functionals*, loc. cit. For essentially Martin's proof of the existence and uniqueness of a polar see also Taylor, A. E., "Additions to the Theory of Polynomials in Normed Linear Spaces," *Tôhoku Math. Jour.*, **44**, 302–318 (1938), where other references are given. See also Van Der Lijn, *Bulletin des Sciences Math.* (1940).

[5] The system (27) in Fréchet differentials is not of the type for which existence theorems were given by Michal and Elconin, loc. cit. For some further results on the matric exponential, the reader is referred to Michal, A. D., *The Total Differential Equation for the Exponential Function in Non-Commutative Normed Linear Rings* (to be published). For some applications of the matric exponential to vibration problems, the reader is referred to Michal, A. D., *Matrix and Tensor Calculus with Applications to Mechanics, Elasticity, and Aeronautics* (GALCIT series of John Wiley and Sons, in press).

[6] The functional of $\xi$ defined by $L(K, \xi) = \xi + \xi K$ is a solvable linear functional of $\xi$ with inverse $M(K, \xi) = \xi + \xi k[K]$. Hence by a known result (Michal and Elconin, loc. cit.) the Fréchet differential of $\xi k[K]$ exists for each $K(x, \xi) \epsilon \Delta$. With this result one can complete the proof of Theorem 5 without much difficulty. This same type of proof can also be used in proving the corresponding result (5) for Volterra integral equations.

# ON THE PROJECTIVE THEORY OF TWO DIMENSIONAL RIEMANN SPACES

BY T. Y. THOMAS

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY

Communicated June 25, 1945

The only significant projective invariant which has been defined in the strict domain of the projective space of paths is the projective curvature tensor discovered by Weyl[1] in 1921. In the following note we show the existence of a simple *vector* invariant of the two-dimensional projective Riemann space (defined by the totality of Riemann metrics which yield the same system of geodesics). The vanishing of this vector constitutes the necessary and sufficient condition for the space to be of constant curvature in the projective sense.

The construction of the above vector invariant follows readily from equations which have already appeared in the literature.[2] Thus if $g_{\alpha\beta}(x)$ and $\bar{g}_{\alpha\beta}(x)$ are the components of metric tensors yielding the same system of geodesics we must have

$$\bar{\Gamma}^{\alpha}_{\beta\gamma} = \Gamma^{\alpha}_{\beta\gamma} + \delta^{\alpha}_{\beta}\phi_{\gamma} + \delta^{\alpha}_{\gamma}\phi_{\beta}, \tag{1}$$

where the $\Gamma$'s and $\bar{\Gamma}$'s are the Christoffel symbols determined by these tensors and the $\phi$'s are the components of a covariant vector. Equations (1) are completely equivalent to the system

$$\bar{g}_{\alpha\beta,\,\gamma} = \bar{g}_{\alpha\gamma}\phi_{\beta} + \bar{g}_{\beta\gamma}\phi_{\alpha} + 2\bar{g}_{\alpha\beta}\phi_{\gamma}, \tag{2}$$

in which the comma denotes covariant differentiation based on the metric defined by the $g_{\alpha\beta}$. Contraction of indices in (1) shows that the vector having the components $\phi_{\alpha}$ is the gradient of a scalar function $\phi$ given by

$$\phi = \frac{1}{2(n+1)} \log\left(\frac{\bar{g}}{g}\right),$$

where $g$ and $\bar{g}$ denote the determinants formed from the components $g_{\alpha\beta}$ and $\bar{g}_{\alpha\beta}$, respectively, and $n$ is the dimensionality of the space. Substitution of the components $\bar{\Gamma}^{\alpha}_{\beta\gamma}$ from (1) into the expression for the components of the ordinary curvature tensor, followed by contraction of indices, leads to the system

$$\bar{B}_{\alpha\beta} = B_{\alpha\beta} + (n-1)[\phi_{\alpha}\phi_{\beta} - \phi_{\alpha\beta}], \tag{3}$$

where the $B_{\alpha\beta}$ and $\bar{B}_{\alpha\beta}$ are the components of the contracted curvature tensors and the $\phi_{\alpha\beta}$ are the components of the second covariant derivative of the scalar $\phi$.

We now suppose $n = 2$. Then the curvature tensor components $B^{\alpha}_{\beta\gamma\delta}$ can be written

$$B^{\alpha}_{\beta\gamma\delta} = g^{\alpha\sigma}B_{\sigma\beta\gamma\delta} = g^{\alpha\sigma}(g_{\sigma\delta}g_{\beta\gamma} - g_{\sigma\gamma}g_{\beta\delta})K = (\delta^{\alpha}_{\delta}g_{\beta\gamma} - \delta^{\alpha}_{\gamma}g_{\beta\delta})K, \quad (4)$$

where $K$ is the Gaussian curvature. Putting $\alpha = \delta$ in these equations and summing we find $B_{\beta\gamma} = Kg_{\beta\gamma}$. Hence (3) becomes

$$\phi_{\alpha\beta} = \phi_{\alpha}\phi_{\beta} + Kg_{\alpha\beta} - \bar{K}\bar{g}_{\alpha\beta}. \quad (5)$$

Now differentiate (5) covariantly with respect to $x^{\gamma}$, interchange the indices $\beta$ and $\gamma$ and subtract. This gives

$$\phi_{\alpha\beta,\,\gamma} - \phi_{\alpha\gamma,\,\beta} = \phi_{\alpha\gamma}\phi_{\beta} - \phi_{\alpha\beta}\phi_{\gamma} + K_{\gamma}g_{\alpha\beta} - K_{\beta}g_{\alpha\gamma} - \bar{K}_{\gamma}\bar{g}_{\alpha\beta} + \bar{K}\bar{g}_{\alpha\gamma} - \bar{K}\bar{g}_{\alpha\beta,\,\gamma} + \bar{K}\bar{g}_{\alpha\gamma,\,\beta}. \quad (6)$$

For the left member of these equations we have

$$\phi_{\alpha\beta,\,\gamma} - \phi_{\alpha\gamma,\,\beta} = -\phi_{\sigma}B^{\sigma}_{\alpha\beta\gamma} = (\phi_{\beta}g_{\alpha\gamma} - \phi_{\gamma}g_{\alpha\beta})K.$$

Making this substitution, and the substitutions (2) and (5) in the right member of (6), the resulting expression is seen to reduce to

$$\bar{K}_{\beta}\bar{g}_{\alpha\gamma} - \bar{K}_{\gamma}\bar{g}_{\alpha\beta} = K_{\beta}g_{\alpha\gamma} - K_{\gamma}g_{\alpha\beta}. \quad (7)$$

It follows from (7) that the quantities $K_{\beta}g_{\alpha\gamma} - K_{\gamma}g_{\alpha\beta}$ are the components of a projective Riemann tensor. Since these quantities are skew symmetric in the indices $\beta$ and $\gamma$ this tensor involves at most two independent components. This suggests that the tensor is equivalent to a vector invariant. To show that such is the case we proceed as follows: make the substitutions

$$g_{11} = gg^{22},\ g_{22} = -gg^{12},\ g_{22} = gg^{11}$$

and similar substitutions in the quantities $\bar{g}_{\alpha\beta}$ in the equations (7). These equations then yield

$$\bar{g}\bar{g}^{\alpha\beta}\bar{K}_{\beta} = gg^{\alpha\beta}K_{\beta}. \quad (8)$$

*Hence the quantities $gg^{\alpha\beta}K_{\beta}$ are the components of a relative projective vector of weight two of the two dimensional Riemann space.* The vanishing of this vector implies $K_{\alpha} = 0$ and conversely. Thus the only spaces whose geodesics are identical with the geodesics of a space of constant curvature are spaces of constant curvature (Beltrami). Moreover if $K = \text{const.}$, the constant $\bar{K}$ can be chosen arbitrarily; this follows readily since under these conditions the system consisting of (2) and (5) is completely integrable. The totality of Riemann spaces of constant curvature having the same geodesics can conveniently be described as a projective Riemann space of constant curvature. Hence the condition for a two-dimensional space to be of constant curvature in the projective sense (projective Rie-

mann space of constant curvature) is the vanishing of the above projective vector invariant.

It follows from (8) that the system of curves defined by

$$\frac{dx^\alpha}{dt} = g^{\alpha\beta}K_\beta \tag{9}$$

has an invariant determination in the two-dimensional projective space, i.e., *the congruence given by* (9) *is independent of the particular metric tensor by which the geodesics of the space are determined.* The family of curves (9) is of especial interest since in general these curves are distinct from the invariant geodesics of the space. It would appear that this invariant family of curves (or the above projective vector invariant) would be of particular significance in the further study of the geometry of the two-dimensional projective Riemann space.

[1] Weyl, H., "Zur Infinitesimalgeometrie: Einordnung der projektiven und der konformen Auffassung," *Göttingen Nachrichten*, 1921, pp. 99–112. A very complicated projective tensor of no stated significance has been constructed by J. M. Thomas, these PROCEEDINGS, **11**, 207–209 (1925).

[2] See, for example, Eisenhart, L. P., *Riemannian Geometry*, Princeton University Press, 1926, pp. 131–135.

---

## *ON CONVERGENCE IN LENGTH*

BY MIRIAM C. AYER

THE OHIO STATE UNIVERSITY AND WELLESLEY COLLEGE

Communicated June 13, 1945

1. The topics discussed in this note have their origin in the theory of arc length. We shall be concerned with triples of equations of the form $x = x(u), y = y(u), z = z(u)$ where each function is defined and continuous on a closed linear interval $\alpha \leq u \leq \beta$. For conciseness we write the triple as $\mathfrak{x} = \mathfrak{x}(u)$, $u \in I$, where $\mathfrak{x}(u)$ is the vector $(x(u), y(u), z(u))$ and $I$ is the interval $\alpha \leq u \leq \beta$. With length defined in the usual way in terms of inscribed polygons, the length of the curve represented by $\mathfrak{x} = \mathfrak{x}(u)$, $u \in I$, is then a function $L(\mathfrak{x})$ of the vector $\mathfrak{x}$.

The following definitions explain the notation and terminology used. The magnitude of a vector $\mathfrak{x}$ is denoted by $|\mathfrak{x}|$. Given any vector function $\mathfrak{x}(u) = (x(u), y(u), z(u))$, $u \in I$, the vector $(x'(u), y'(u), z'(u))$, defined wherever these derivatives exist, is written as $\mathfrak{x}'(u)$ or, more briefly, $\mathfrak{x}'$. A vector $\mathfrak{x}(u)$, $u \in I$, is said to be continuous if and only if each component $x(u), y(u), z(u)$ is continuous on $I$. A vector $\mathfrak{x}(u)$, $u \in I$, is said to be of bounded variation (briefly, $\mathfrak{x}$ is BV on $I$) if and only if each component

$x(u)$, $y(u)$, $z(u)$ is of bounded variation on $I$. A vector $\mathfrak{x}(u)$, $u \in I$, is said to be absolutely continuous (briefly, AC) if and only if each component $x(u)$, $y(u)$, $z(u)$ is absolutely continuous on $I$. The functional $L(\mathfrak{x})$ is defined explicitly as follows: $L(\mathfrak{x}) = lub\Sigma|\mathfrak{x}(u'') - \mathfrak{x}(u')|$, where the least upper bound is taken over all finite subdivisions of $I$ into non-overlapping intervals $u' \leq u \leq u''$. For vectors of the special form $\mathfrak{x}(u) = (f(u), 0, 0)$ the length reduces to the total variation $V(f)$ of $f(u)$ on $I$.

We state here some well-known facts about arc length.[6] $L(\mathfrak{x})$ is finite if and only if $\mathfrak{x}$ is BV on $I$. If $L(\mathfrak{x})$ is finite, then $\mathfrak{x}'(u)$ exists a.e. (almost everywhere) on $I$, $|\mathfrak{x}'(u)|$ is summable on $I$, and $L(\mathfrak{x}) \geq \int_I |\mathfrak{x}'|$. Equality holds in this relation if and only if $\mathfrak{x}$ is AC on $I$.

Uniform convergence on $I$ of a sequence $f_n(u)$, $u \in I$, $n = 0, 1, 2, \ldots$ is written as $f_n \to f_0\{\mathfrak{U}\}$. Similarly, $f_n \to f_0\{\mathfrak{M}\}$ on $E$ denotes convergence in measure of $f_n$ to $f_0$ on the set $E$. A sequence of vectors $\mathfrak{x}_n(u) = (x_n(u), y_n(u), z_n(u))$, $u \in I$, $n = 0, 1, 2, \ldots$ is said to converge uniformly (briefly, $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$) if and only if $x_n(u)$, $y_n(u)$, and $z_n(u)$ converge uniformly on $I$ to $x_0(u)$, $y_0(u)$ and $z_0(u)$, respectively. It is well known[1, 2, 5] that if $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$, $\mathfrak{x}_n$ BV and continuous for $n = 0, 1, 2, \ldots$, then $\liminf L(\mathfrak{x}_n) \geq L(\mathfrak{x}_0)$. When $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$ and $L(\mathfrak{x}_n) \to L(\mathfrak{x}_0)$, then we shall say that $\mathfrak{x}_n$ converges to $\mathfrak{x}_0$ in length, in symbols: $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$. Hereafter all vectors $\mathfrak{x}(u)$, $u \in I$, will be assumed to be both BV and continuous on $I$.

In the literature on this subject[2, 5] previous results relate convergence in length with the following: (i) convergence in direction, i.e., convergence in some appropriate sense of $\mathfrak{x}_n'(u)$ to $\mathfrak{x}_0'(u)$, and (ii) some kind of convergence of each of the sequences $x_n(u)$, $y_n(u)$ and $z_n(u)$, $n = 0, 1, 2, \ldots$. In these results there is a sharp distinction between the parametric and the non-parametric cases, the latter being the special case in which $\mathfrak{x}_n(u) = (u, y_n(u), z_n(u))$ for $n = 0, 1, 2, \ldots$. The non-parametric case appears to be somewhat simpler than the more general parametric case and certain results valid in the non-parametric case seem to admit of no extension to the parametric case. The purpose of this paper is to make further contributions to the theory along these lines and, in particular, to narrow the gap between the parametric and non-parametric cases.

2. In previous literature[2, 5] convergence in variation and strong convergence in variation are considered along with convergence in length. For sequences of vectors, however, we shall use strong convergence in length instead of strong convergence in variation. (These two conceptions are equivalent, and for vectors of the special form $\mathfrak{x}_n = (f_n(u), 0, 0)$ strong convergence in length of $\mathfrak{x}_n$ reduces to strong convergence in variation of the scalars $f_n(u)$.) We state the definitions of the types of convergence used in this paper. A sequence $f_n(u)$, $u \in I$, $f_n$ BV and continuous on $I$, $n = 0, 1, 2, \ldots$, is said to converge in variation, briefly $f_n \to f_0\{\mathfrak{V}\}$, if and only if $f_n \to f_0\{\mathfrak{U}\}$ and $V(f_n) \to V(f_0)$. A sequence of vectors $\mathfrak{x}_n(u) =$

$(x_n(u), y_n(u), z_n(u))$, $n = 0, 1, 2, \ldots$, is said to converge in variation, briefly $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$, if and only if $x_n \to x_0\{\mathfrak{U}\}$, $y_n \to y_0\{\mathfrak{U}\}$, and $z_n \to z_0\{\mathfrak{U}\}$. A sequence of vectors $\mathfrak{x}_n(u)$, $\mathfrak{x}_n$ BV and continuous on $I$, $n = 0, 1, 2, \ldots$, is said to converge strongly in length, briefly $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{SL}\}$, if and only if $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$ and $L(\mathfrak{x}_n - \mathfrak{x}_0) \to 0$.

It is easily verified that strong convergence in length implies convergence in length, but not conversely. Moreover it is well known[2, 5] that convergence in variation is necessary but not sufficient for convergence in length. Results stated in later sections will clarify further the relations among these convergence types.

3. If $\mathfrak{x}(u)$ is any given vector continuous and BV on $I$, each subinterval $\Delta \subset I$ determines an arc of length $L(\Delta, \mathfrak{x})$. In this notation the quantity previously denoted by $L(\mathfrak{x})$ is written as $L(I, \mathfrak{x})$. The interval function $L(\Delta, \mathfrak{x})$ is non-negative, continuous and additive, and therefore can be extended to a c.a. (completely additive) function $L(E, \mathfrak{x})$ of Borel sets $E$ in $I$. By the general theory of such set functions we then have the Lebesgue decomposition[6] $L(E, \mathfrak{x}) = L_a(E, \mathfrak{x}) + L_s(E, \mathfrak{x})$, where $L_a(E, \mathfrak{x}) = \int_E |\mathfrak{x}'|$ is a non-negative, c.a., AC function of Borel sets and $L_s(E, \mathfrak{x})$ is a non-negative, c.a., singular function of Borel sets. A BV function $f(u)$, $u \in I$, is expressible as the sum of an AC function and a singular function, this decomposition being univocally determined if we agree that the singular part vanish at the left end point of $I$. This unique decomposition will be called the normal Lebesgue decomposition of $f$ and will be denoted by $f = f_a + f_s$, where $f_a$ is AC and $f_s$ singular. In terms of the normal Lebesgue decompositions of the components $x(u)$, $y(u)$, $z(u)$ we then introduce the normal Lebesgue decomposition of the vector $\mathfrak{x}$ as $\mathfrak{x}_a + \mathfrak{x}_s$, where $\mathfrak{x}_a = (x_a, y_a, z_a)$ and $\mathfrak{x}_s = (x_s, y_s, z_s)$. The vectors $\mathfrak{x}_a$ and $\mathfrak{x}_s$ give rise, in turn, to non-negative, c.a. functions of Borel sets $L(E, \mathfrak{x}_a)$ and $L(E, \mathfrak{x}_s)$, respectively. The two Lebesgue decompositions $L(E, \mathfrak{x}) = L_a(E, \mathfrak{x}) + L_s(E, \mathfrak{x})$ and $\mathfrak{x}(u) = \mathfrak{x}_a(u) + \mathfrak{x}_s(u)$ are related in the following simple way.

LEMMA: *If $\mathfrak{x}(u)$ is BV and continuous on $I$, then $L_a(E, \mathfrak{x}) = L(E, \mathfrak{x}_a)$ and $L_s(E, \mathfrak{x}) = L(E, \mathfrak{x}_s)$ for every Borel set $E \subset I$.*

The Lemma of the Lebesgue decomposition leads to a number of new results and makes it possible to simplify proofs of several known results. By way of illustration we state some of the implications of the lemma.

(*a*) If $\mathfrak{x}_n(u)$, $u \in I$, $n = 0, 1, 2, \ldots$, is a sequence of continuous, BV vectors, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{SL}\}$ if and only if $\mathfrak{x}_{na} \to \mathfrak{x}_{0a}\{\mathfrak{SL}\}$ and $\mathfrak{x}_{ns} \to \mathfrak{x}_{0s}\{\mathfrak{SL}\}$.

(*b*) If $\mathfrak{x}_1(u)$ and $\mathfrak{x}_2(u)$ are BV and continuous on $I$, then $L(\mathfrak{x}_1 + \mathfrak{x}_2) = L(\mathfrak{x}_1) + L(\mathfrak{x}_2)$ if and only if $L(\mathfrak{x}_{1a} + \mathfrak{x}_{2a}) = L(\mathfrak{x}_{1a}) + L(\mathfrak{x}_{2a})$ and $L(\mathfrak{x}_{1s} + \mathfrak{x}_{2s}) = L(\mathfrak{x}_{1s}) + L(\mathfrak{x}_{2s})$.

(*c*) If $\mathfrak{x} = (u, f(u), 0)$, $f(u)$ continuous and BV on $I$, then $L(\mathfrak{x}) \leq (\beta - \alpha) + V(f)$, equality holding if and only if $f(u)$ is singular.

(*d*) If $\mathfrak{x}_n$ is BV and continuous for $n = 1, 2, \ldots$ and $\mathfrak{x}_0$ is AC, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{SL}\}$ if and only if $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$ and $|\mathfrak{x}_n' - \mathfrak{x}_0'| \to 0\{\mathfrak{M}\}$ on $I$.

(*e*) If $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$, $\mathfrak{x}_n$ AC for $n = 1, 2, \ldots$, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{SL}\}$ if and only if $\int_I |\mathfrak{x}_n' - \mathfrak{x}_0'| \to 0$.

(*f*) If $\mathfrak{x} = \mathfrak{x}_n(u)$, $n = 1, 2, \ldots$, represent polygons inscribed in the curve represented by $\mathfrak{x} = \mathfrak{x}_0(u)$, where $\mathfrak{x}_0$ is BV and continuous, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{SL}\}$ if and only if $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$ and $\mathfrak{x}_0$ is AC.

Let us add a remark concerning statement (*c*). In the special case where $f(u)$ is monotone and singular, (*c*) implies that $L(\mathfrak{x}) = (\beta - \alpha) + |f(\beta) - f(\alpha)|$. That is, the length of the curve $y = f(x)$ is then equal to the sum of its projections on the $x$ and $y$ axes. This can be regarded as a generalization of the well-known fact[2] that the curve determined by the Cantor function $y = f(x)$, $0 \leq x \leq 1$, has length 2.

4. The Steiner inequality is an important tool in previous literature.[3, 4] This inequality states that $L([\mathfrak{x}_1 + \mathfrak{x}_2]/2) \leq [L(\mathfrak{x}_1) + L(\mathfrak{x}_2)]/2$ for any two vectors $\mathfrak{x}_1(u)$, $\mathfrak{x}_2(u)$ which are BV and continuous on $I$. Although the denominator 2 is essential if we confine our attention to vectors in the non-parametric form, it is generally more convenient to omit it.

It is of interest to find conditions under which the sign of equality will hold. (See §2 (*b*).) In the non-parametric case we have the well-known simple result: If $\mathfrak{x}_1 = (u, f_1(u), 0)$, $\mathfrak{x}_2 = (u, f_2(u), 0)$, where $f_1(u)$ and $f_2(u)$ have continuous derivatives on $I$, then equality holds in the Steiner inequality if and only if $f_1(u)$ and $f_2(u)$ differ by a constant. That is, for equality to hold two such curves must be parallel. The following statement is a generalization of this.

(*a*) If $\mathfrak{x}_1(u)$ and $\mathfrak{x}_2(u)$ are AC on $I$, then $L(\mathfrak{x}_1 + \mathfrak{x}_2) = L(\mathfrak{x}_1) + L(\mathfrak{x}_2)$ if and only if $\mathfrak{x}_1'/|\mathfrak{x}_1'| = \mathfrak{x}_2'/|\mathfrak{x}_2'|$ a.e.$_2'$ on $E = \underset{u}{E}[u \in I, |\mathfrak{x}_1'| > 0, |\mathfrak{x}_2'| > 0]$.

In the literature[3, 4] there are various generalizations involving a study of finding conditions under which the sign of equality will hold approximately. That work is concerned solely with the non-parametric case. The following statements may be considered as extensions of those results to the general case.

(*b*) If $\mathfrak{x}_1(u)$ and $\mathfrak{x}_2(u)$ are BV and continuous on $I$, then each of the following hold:

(i) $[\int_I \{|\mathfrak{x}_1'| \cdot |\mathfrak{x}_2'| - \mathfrak{x}_1'\mathfrak{x}_2'\}^{1/2}]^2 \leq [L(\mathfrak{x}_1) + L(\mathfrak{x}_2)] \cdot [L(\mathfrak{x}_1) + L(\mathfrak{x}_2) - L(\mathfrak{x}_1 + \mathfrak{x}_2)]$

(ii) $[\int_I \{(\mathfrak{x}_1')^2(\mathfrak{x}_2')^2 - (\mathfrak{x}_1'\mathfrak{x}_2')^2\}^{1/4}]^4 \leq {}^1/_2[L(\mathfrak{x}_1) + L(\mathfrak{x}_2)]^3 \cdot [L(\mathfrak{x}_1) + L(\mathfrak{x}_2) - L(\mathfrak{x}_1 + \mathfrak{x}_2)]$

In the non-parametric case where $\mathfrak{x}_1 = (u, y_1(u), z_1(u))$ and $\mathfrak{x}_2 = (u, y_2(u), z_2(u))$ we have $\mathfrak{x}_1' = (1, y_1'(u), z_1'(u))$ and $\mathfrak{x}_2' = (1, y_2'(u), z_2'(u))$. In this case an elementary discussion yields the relation $|\mathfrak{x}_1' - \mathfrak{x}_1'|^2 \leq (\mathfrak{x}_1')^2(\mathfrak{x}_2')^2 - (\mathfrak{x}_1'\mathfrak{x}_2')^2$. From part (ii) of (*b*) we then obtain the following result.

(*c*) If $\mathfrak{x}_1 = (u, y_1(u), z_1(u))$ and $\mathfrak{x}_2 = (u, y_2(u), z_2(u)$ are BV and continuous on $I$, then $\{\int_I |\mathfrak{x}_1' - \mathfrak{x}_2'|^{1/2}\}^4 \leq \{L(\mathfrak{x}_1) + L(\mathfrak{x}_2)\}^2 \cdot \{[L(\mathfrak{x}_1) + L(\mathfrak{x}_2)]/2 -$

$L([\mathfrak{x}_1 + \mathfrak{x}_2]/2)\}$. This inequality is the analogue of a known inequality[4] involving surfaces. Results known for the non-parametric case now appear as corollaries of new results for the parametric case.

The following statements are corollaries of inequalities (*b*) and (*c*).

(*d*) If $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$, then $|\mathfrak{x}_n'| \cdot |\mathfrak{x}_0'| - \mathfrak{x}_n'\mathfrak{x}_0' \to 0\{\mathfrak{M}\}$ on $I$ and $(\mathfrak{x}_n')^2(\mathfrak{x}_0')^2 - (\mathfrak{x}_n'\mathfrak{x}_0')^2 \to 0\{\mathfrak{M}\}$ on $I$.

(*e*) If $\mathfrak{x}_n = (u, y_n(u), z_n(u))$ for $n = 0, 1, 2, \ldots$ and if $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$, then $|\mathfrak{x}_n' - \mathfrak{x}_0'| \to 0\{\mathfrak{M}\}$ on $I$.

5. Statement (*e*) of §4 is an example of a case where convergence in length implies some kind of convergence of the derivatives. That statement is concerned with the non-parametric case and we would like to find its analogue for the parametric case. Let $t_n$ denote the unit tangent vector $\mathfrak{x}_n'/|\mathfrak{x}_n'|$, defined wherever $|\mathfrak{x}_n'| > 0$, $n = 0, 1, 2, \ldots$. Results concerning the behavior of $t_n$ are obtained from (*d*) of § 4.

(*a*) If $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$, then $\int_E |t_n - t_0| \to 0$, where $E = \underset{u}{E}[u \in I, |\mathfrak{x}_0'| > 0, \liminf |\mathfrak{x}_n'| > 0]$.

(*b*) If $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$ and $\mathfrak{x}_n = (u, y_n(u), z_n(u))$ for $n = 0, 1, 2, \ldots$, then $\int_I |t_n - t_0| \to 0$.

In a sense, the preceding statements can be considered as relating convergence in length to convergence in direction. Although the converses are not true, the following special case is of interest.

(*c*) If $\mathfrak{x}_n = (u, y_n(u), z_n(u))$, $u \in I$, $n = 0, 1, 2, \ldots$, and $\mathfrak{x}_0$ is AC, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$ if and only if $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$ and $|\mathfrak{x}_n' - \mathfrak{x}_0'| \to 0\ \{\mathfrak{M}\}$ on $I$.

(*d*) If $\mathfrak{x}_n = (u, y_n(u), z_n(u))$, $u \in I$, $n = 0, 1, 2, \ldots$, and $\mathfrak{x}_0$ is AC, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$ if and only if $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$ and $\int_I |t_n - t_0| \to 0$.

6. There is apparently a lack of finality and completeness in the theory of convergence in length. The reason for this is that some of the simple properties of strong convergence in length do not generally hold for convergence in length. In special cases where suitable additional conditions are imposed on the vector functions complete characterizations are possible. Statements (*c*) and (*d*) of § 5 illustrate such a case.

A sequence $\mathfrak{x}_n(u) = (x_n(u), y_n(u), z_n(u))$, $u \in I$, $n = 0, 1, 2, \ldots$ is said to be uniformly AC if and only if $x_n(u)$, $y_n(u)$, $z_n(u)$, $n = 0, 1, 2, \ldots$ are equi-absolutely continuous on $I$. In the non-parametric case this is a necessary condition for a sequence of AC vectors to converge in length.

(*a*) If $\mathfrak{x}_n(u) = (u, y_n(u), z_n(u))$, $u \in I$, $\mathfrak{x}_n$ AC, $n = 0, 1, 2, \ldots$, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$ if and only if the vectors $\mathfrak{x}_n$, $n = 0, 1, 2, \ldots$ are uniformly AC and $|\mathfrak{x}_n' - \mathfrak{x}_0'| \to 0\{\mathfrak{M}\}$ on $I$.

In the non-parametric case the absolute continuity of $\mathfrak{x}_0$ is sufficient to make convergence in length equivalent to strong convergence in length. Obtaining characterization theorems for the cases just mentioned is therefore a simple problem. In the following result, however, strong convergence in length is not implied.

(*b*) If $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{U}\}$, $\mathfrak{x}_n$ uniformly AC on $I$ for $n = 0, 1, 2, \ldots$, then $\mathfrak{x}_n \to \mathfrak{x}_0\{\mathfrak{L}\}$ if and only if

$$\text{(i)} \quad |\mathfrak{x}_n'| \cdot |\mathfrak{x}_0'| - \mathfrak{x}_n' \mathfrak{x}_0' \to 0\{\mathfrak{M}\} \text{ on } I$$
$$\text{(ii)} \quad |\mathfrak{x}_n'| \to 0\{\mathfrak{M}\} \text{ on } E = \underset{u}{E}[u \in I, |\mathfrak{x}_0'| = 0].$$

7. For vectors of the special form $\mathfrak{x}_n = (f_n(u), 0, 0)$ the results of §§ 3–6 yield as corollaries corresponding statements on convergence in variation. We state a few of these.

(*a*) If $f_1(u)$ and $f_2(u)$ are $BV$ and continuous on $I$, then $V(f_1 + f_2) = V(f_1) + V(f_2)$ if and only if $f_1' f_2' \geq 0$ a.e. on $I$ and $V(f_{1s} + f_{2s}) = V(f_{1s}) + V(f_{2s})$.

(*b*) If $f_n \to f_0\{\mathfrak{U}\}$, then $|f_n f_0'| - f_n f_0' \to 0\{\mathfrak{M}\}$ on $I$.

(*c*) If $f_n \to f_0\{\mathfrak{U}\}$, $f_n$ equi-absolutely continuous on $I$ for $n = 0, 1, 2, \ldots$, then $f_n \to f_0\{\mathfrak{U}\}$ if and only if

$$\text{(i)} \quad |f_n' f_0'| - f_n' f_0' \to 0\{\mathfrak{M}\} \text{ on } I$$
$$\text{(ii)} \quad f_n' \to 0\{\mathfrak{M}\} \text{ on } E = \underset{u}{E}[u \in I, f_0' = 0].$$

[1] Adams, C. R., and Clarkson, J. A., *Bull. Am. Math. Soc.*, **40**, 413–417 (1934).
[2] Adams, C. R., and Lewy, H.; *Duke Math. Jour.*, **1**, no. 1, 19–26 (1935).
[3] McShane, E. J., *Ann. Math.*, (2) **33**, 125–138 (1932).
[4] Radó, T., and Reichelderfer, P., *Bull. Am. Math. Soc.*, **47**, no. 2, 102–108 (1941).
[5] Radó, T., and Reichelderfer, P., *Duke Math. Jour.*, **9**, no. 3, 527–565 (1942).
[6] Saks, S., *Theory of the Integral*, Warszawa-Lwów, 1937.